Neuroscience Letters 471 (2010) 100–103

Contents lists available at ScienceDirect

Neuroscience Letters

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

Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading

Martijn Baart, Jean Vroomen*

Department of Medical Psychology and Neuropsychology, Tilburg University, P.O. Box 90153, Warandelaan 2, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 30 April 2009
Received in revised form
14 December 2009
Accepted 9 January 2010

Keywords:
Ambiguous lipread speech
Recalibration
Selective speech adaptation
Bi-directional integration

ABSTRACT

It is well known that visual information derived from mouth movements (i.e., lipreading) can have profound effects on auditory speech identification (e.g. the McGurk-effect [16]). Here we examined the reverse phenomenon, namely whether auditory speech affects lipreading. We report that speech sounds dubbed onto lipread speech affect immediate identification of lipread tokens. This effect likely reflects genuine cross-modal integration of sensory signals and not just a simple response bias because we also observed adaptive shifts in visual identification of the ambiguous lipread tokens after exposure to incongruent audiovisual adapter stimuli. Presumably, listeners had learned to label the lipread stimulus in accordance with the sound, thus demonstrating that the interaction between hearing and lipreading is genuinely bi-directional.

The question of how sensory modalities cooperate in forming a coherent representation of the environment is the focus of much current work. A particularly elucidating example is the interaction between hearing and seeing speech (here referred to as lipreading). In one of the more spectacular cases, listeners report to 'hear' /da/ when in fact, auditory /ba/ is dubbed onto lipread /ga/, the McGurk-effect [16]. Numerous studies have explored the brain mechanisms underlying this phenomenon. Some have reported that visual speech may affect auditory processing as early as the auditory cortex [8,10,17,19,24]. The interaction has been found to occur between 150 and 250 ms using the mismatch negativity paradigm [10,17,24], while others have reported that as early as 100 ms, the auditory N1 component is attenuated and speeded up when auditory speech is accompanied by lipread information [4,29], possibly because visual speech predicts when a sound is going to occur [26,33].

Notably though, to date it is not known whether auditory speech also affects visual processing of lipread speech. This is surprising because bi-directional effects have been reported in other cross-modal illusions. For example, in the 'ventriloquist illusion', the apparent location of a sound is displaced towards a simultaneously presented and spatially misaligned light [2]. The reverse phenomenon, namely that the apparent location of a visual target is shifted towards an auditory displaced distracter, has also been reported [20], although the effect is admittedly small because the more reliable information source, – for space vision – is dominant and thus less susceptible to cross-modal biases [11,14].

Here, we sought to show that identification of lipread stimuli is affected by speech sounds. For that purpose, we created a 7-point continuum of visual stimuli in between /omso/ and /onso/. Participants were instructed to lipread these stimuli and press an 'm'- or 'n'-key upon lipreading /omso/ or /onso/, respectively (a visual 2AFC-task), while trying to ignore /omso/ or /onso/ sounds that were dubbed onto the videos. Despite instructions to ignore the sound, we expected the sound to shift the visual identification function of the lipread stimuli, so more 'n'-responses upon hearing /onso/ rather than /omso/.

Twelve native speakers of Dutch (mean age = 23) with normal hearing and normal or corrected-to-normal vision participated (mean age = 21) after giving written informed consent. The experiment was conducted in accordance with the Declaration of Helsinki.

Stimulus creation started with two videos of the full face of a male speaker pronouncing the pseudo-words /omso/ and /onso/ as previously used by Tuomainen et al. [27]. The head, nose, and eye position of the speaker were well aligned, so fusion of the stimuli could be accomplished by adjusting the overall opacity rather than applying a morphing technique with landmarks on the face. The lipread /m/ and /n/ belong to different 'viseme' classes [36], and are thus relatively easy to visually discriminate. To create a continuum in between the two recordings, videos were first converted into bitmap sequences (29.97 f/s) matched for total duration (45 bitmaps; 1500 ms) and for onset and offset of the articulatory gesture (at 567 and 1367 ms, respectively). Each individual bitmap of the /omso/ sequence was fused with the corresponding /onso/ bitmap by adding the two bitmaps in different relative proportions

* Corresponding author. Tel.: +31 13 4662394.
E-mail address: j.vroomen@uvt.nl (J. Vroomen).
URL: http://spitswww.uvt.nl/ vroomen (J. Vroomen).

0304-3940/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2010.01.019

to each other. Seven bitmap sequences were created by varying the relative proportion from 0 to 100% for the most /omso/-like stimulus, through 15–85%, 29–81%, 43–57%, 58–42%, 72–28%, to 90–10% for the most /onso/-like stimulus. Each of the seven thus created videos (14.9 (*H*) by 18.8 (*W*) cm in size) looked natural without any noticeable jitter or fading. The natural timing between the audio and video was preserved by relying on a custom-made programme that displayed the bitmap sequence and played the sound by trigger, rather than a standard PC-based video-player whose timing was considered to be too unreliable.

Participants were tested individually in a sound attenuated and dimly lit booth and were seated at approximately 70 cm from a 17-in. CRT screen. The audio was presented at 63 dBa at ear-level via two regular loudspeakers placed left and right of the monitor. The seven lipread tokens of the continuum were delivered in combination with auditory /omso/ and /onso/, and in a silent condition. Each of the 21 stimuli was delivered 20 times (420 trials in total) in four blocks of 105 randomly presented trials. Participants judged whether the visual stimulus was /omso/ or /onso/ by pressing the corresponding 'm'- or 'n'-key. The next trial started 750 ms after a response was detected. Prior to testing, participants received a short practice session (12 trials) in which they were shown the two extreme tokens of the lipread continuum combined with auditory /omso/, /onso/, or silence. It was stressed that participants had to rely on lipreading rather than sound, as the sound did not predict in any sense what the visual stimulus would look like.

For each participant, the proportion of 'n'-responses was determined as a function of the lipread token. The group-averaged data are presented in Fig. 1. As is clearly visible, three rather sharp S-shaped visual identification functions were obtained. The 50% cross-over point of the curves was near the middle of the continuum and the extremes were almost entirely judged as /omso/ or /onso/. This demonstrates that our continuum was adequately created. Most importantly, the dubbing of a sound onto the videos shifted the visual identification functions in the predicted direction, so more 'n'-responses if /onso/ was dubbed onto the video rather than /omso/. A 3 (A_{onso}, A_{omso}, or V-only) × 7 (lipread token) ANOVA on the proportion of 'n'-responses showed a main effect of lipread token ($F(6,66) = 180.73$, $p < .001$) because – unsurprisingly – there were more 'n'-responses if the lipread videos contained a larger portion of the original /onso/-video. Most importantly, there was a main effect of sound ($F(2,22) = 4.61$, $p < .022$) because there were more 'n'-responses if auditory /onso/ was dubbed onto the video rather than /omso/. The interaction was also significant ($F(12,132) = 2.08$, $p < .023$). Separate paired *t*-tests confirmed that there were more 'n'-responses on the first, second, third, and fifth lipread token if that token was combined with auditory /onso/ rather than /omso/ (all p's$_{one\text{-}tailed} < .05$).

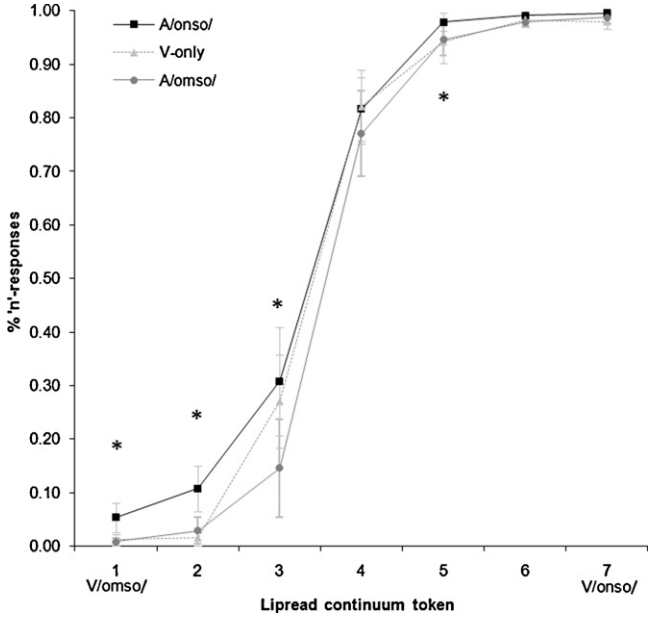

Fig. 1. The proportion of 'n'-responses as a function of the lipread token of the continuum, separately for the visual-only condition, and when combined with auditory /onso/ or /omso/. Significant differences between individual lipread tokens combined with auditory /onso/ versus /omso/ are denoted by an asterisk. Error bars represent one standard error of the mean.

These data thus clearly demonstrate that a speech sound does indeed affect lipreading. The question posed in the introduction, namely whether the cross-modal interaction between speech and lipreading is bi-directional can thus, as a first approximation, be answered affirmatively. However, a critical issue is to determine the processing stage at which this effect occurs. At least two possibilities are available. On the one hand, it may be that the auditory-induced shift is reflecting a truly perceptual effect of sound on vision. Alternatively, though, it might also reflect a response strategy of the participant who, whenever unsure about the visual target, relied on the sound that was heard, despite instructions to ignore that sound.

To further examine this, we conducted another experiment in which we measured aftereffects using an exposure-test paradigm as introduced by Bertelson et al. [3]. In that study, it was reported that if an ambiguous sound halfway between /b/ and /d/ was dubbed onto lipread /b/ (rather than /d/), participants were more likely to categorize the initially ambiguous sound as /b/ when tested later in an auditory-only speech identification test. Presumably, listeners had *learned* to label the ambiguous sound in accord with the lipread information (i.e., phonetic recalibration). This finding was taken as a particularly clear example that lipreading affects speech identification beyond the level of simple response biases. This finding has been replicated ever since in many other studies (e.g. [28,31,32,34,35]). Here, we tested the reverse situation, namely whether a sound would induce a longer-lasting change about the interpretation of an initially ambiguous lipread stimulus. To ensure that this was not due to response priming (i.e., respond /onso/ during test if /onso/ was presented in foregoing exposure phase) we included, as in Bertelson et al. [3], a control condition with stimuli that were not expected to induce recalibration. For that purpose, we used the extreme video tokens /omso/ or /onso/ with the congruent sounds (VmAm and VnAn) dubbed onto it. These stimuli were not expected to induce recalibration because there is no deviance between sight and sound that supposedly drives recalibration. However, the unambiguous nature of the lipread stimuli might possibly cause a contrastive aftereffect (in the auditory domain known as 'selective speech adaptation') as has been demonstrated before for auditory speech [22,23], color, curvature [13], or motion [1], possibly reflecting a 'fatigue' of some hypothetical feature detectors. With unambiguous audiovisual exposure stimuli, one might thus expect *fewer* 'n'-responses after exposure to VnAn than VmAm, an effect in the opposite direction of recalibration.

Twenty-two new native speakers of Dutch with normal hearing and normal or corrected-to-normal vision participated (mean age = 21). A pre-test was used to determine the most ambiguous lipread token for each participant. Each video of the continuum was delivered 16 times in random order and participants indicated whether they saw /omso/ or /onso/. The video closest to the individually determined 50% cross-over point was taken as the perceptually most ambiguous video (henceforth V?). During adaptation proper, participants were repeatedly exposed to a short block of audiovisual adapter stimuli and then tested on lipreading. Each exposure block contained eight consecutive presentations

(ISI = 500 ms) of one of the four audiovisual adapter stimuli V?An or V?Am (to induce recalibration), or VnAn or VmAm (to induce selective adaptation). Exposure was immediately followed by a lipreading test consisting of three different videos presented twice in random order (six test trials in total). The three videos were V?, its immediate 'omso-like' neighbour on the continuum V? − 1, and its immediate 'onso-like' neighbour V? + 1. During the test, participants indicated whether they saw the speaker pronounce /onso/ or /omso/ by pressing a corresponding key.

There were 32 exposure-test blocks in total (8 for each of the 4 adapters), delivered in pseudo-random order. At the end of the experiment, participants were also asked to rate the /omso/–/onso/ quality of the lipread part of the audiovisual adapter stimuli on a seven point Likert-scale with '1' representing a clear visual /omso/ and '7' a clear visual /onso/. Each of the four adapters was presented six times (ISI = 900 ms) in pseudo-random order.

The number of 'n'-responses was determined for each participant (see Fig. 2 for the group averages) and these data were submitted to a 2 (ambiguous or unambiguous lipread exposure) × 2 (adapter sound /omso/ or /onso/) × 3 (lipread test-token) overall ANOVA. There was a main effect of ambiguity of the lipread adapter ($F(1,21) = 5.86$, $p < .025$) because there were somewhat more 'n'-responses after exposure to the ambiguous adapters than the unambiguous ones. The main effect of the lipread test-token ($F(2,42) = 116.42$, $p < .001$) indicated that there were more

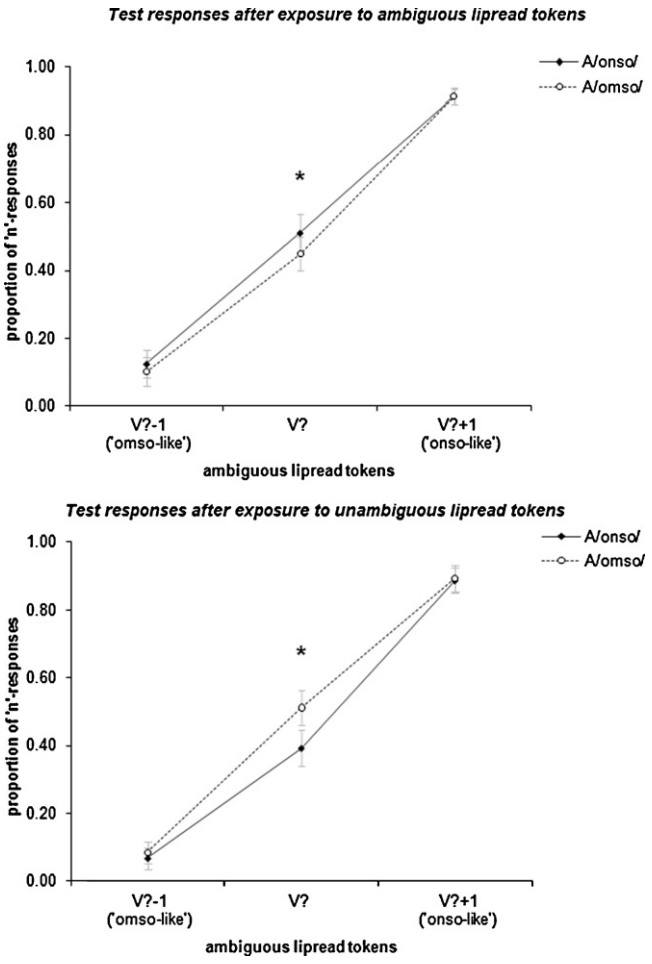

Fig. 2. The mean proportion of 'n'-responses on lipread test-tokens after exposure to ambiguous lipread adapters (AnV? and AmV?; upper panel), and unambiguous lipread adapters (AnVn and AmVm; lower panel). Significant differences between test-tokens preceded by exposure to auditory /onso/ versus /omso/ are denoted by an asterisk. Error bars represent one standard error of the mean.

Table 1
Mean proportion of 'n'-responses and the corresponding aftereffect after exposure to audiovisual adapters with ambiguous and unambiguous lipread videos.

Lipread information	Auditory information		Aftereffect
	/onso/	/omso/	
Ambiguous (V?)	.52	.49	.03
Non-ambiguous (Vn or Vm)	.45	.50	−.05

'n'-responses for the more 'onso-like' token (V? + 1) than for V? and V? − 1. Most importantly, there was an interaction between ambiguity of the lipread adapter and identity of the adapter sound ($F(1,21) = 12.55$, $p < .002$) indicating that there were *more* 'n'-responses after exposure to V?An than V?Am (recalibration), but *fewer* 'n'-responses after exposure to VnAn than VmAm (selective adaptation).

To isolate these effects, aftereffects were calculated analogous to previous studies by subtracting the proportion of 'n'-responses after exposure to auditory /omso/ from /onso/, thereby pooling over the three test-tokens (see Table 1). Separate *t*-tests showed that, in total, there were 3% *more* 'n'-responses after exposure to V?An than V?Am, $t(21) = 2.10$, $p_{\text{one-tailed}} < .024$, while there were 5% *less* 'n'-responses after exposure to VnAn than VmAm, $t(21) = 2.66$, $p_{\text{one-tailed}} < .008$. Fig. 2 shows that the aftereffect was mainly restricted to the most ambiguous token V?. A separate *t*-test isolating performance on the ambiguous V? token showed that there were 6% more 'n'-responses after exposure to V?An than V?Am, $t(21) = 1.91$, $p_{\text{one-tailed}} < .035$, while there were 12% less 'n'-responses after exposure to VnAn than VmAm, $t(21) = 3.39$, $p_{\text{one-tailed}} < .002$. Thus, as predicted, a learning effect was observed if the audiovisual adapter contained an ambiguous lipread token, and a contrast effect if it contained an unambiguous lipread token.

The goodness ratings about the visual part of the audiovisual adapters further confirmed that the ambiguous token V? was rated more 'onso'-like if combined with auditory /onso/ rather than /omso/ (4.51 versus 3.00 on a 7-point scale, respectively, $t(21) = 3.62$, $p_{\text{one-tailed}} < .001$). We also tested the possibility that participants who showed bigger learning effect were also more influenced by the sound of the adapter. The effect of the sound was calculated by taking the difference between the goodness rating of V?An and V?Am, and this difference indeed correlated with the size of the recalibration effect ($r = .39$, $p_{\text{one-tailed}} < .038$). Participants who were strongly affected by the sound thus displayed larger recalibration effects at test.

The present study thus clearly demonstrated that an ambiguous lipread stimulus between /m/ and /n/ is more likely labelled as 'n' if a /n/-sound is dubbed onto it rather than /m/. This immediate effect is not due to a response bias only because we also observed a longer-lasting learning effect: that is, the ambiguous video was labelled more likely as 'n' if in a *preceding adapter phase* an /n/-sound was dubbed onto it rather than /m/. Presumably, exposure to the audiovisual adapter resulted in an endurable adjustment of the boundary of the ambiguous lipread token that – in later testing – was still observable as an aftereffect. For ambiguous lipread tokens, participants thus adjust the phoneme boundary such that the conflict between heard and lipread information is reduced. These findings are in close correspondence with previous reports on phonetic adjustments in auditory speech (e.g. [3,28,31,34,35]), thus indicating that similar mechanisms underlie auditory and lipread recalibration. Moreover, simple response priming (e.g. respond 'n' at test if previously exposed to /onso/) can also be excluded as a mechanism that accounts for these effects because unambiguous and audiovisual congruent adapters produced *contrastive* aftereffects. These visual contrast-effects have been demonstrated before for auditory speech, color, curvature, and so forth, but here we provide the first demonstration of their occurrence for lipread speech.

What might be the functional reason that there is an interaction between seeing and hearing speech? At least two relevant notions have appeared in the literature. The first is that it is 'ecologically' useful to consult more than one source, primarily because different sense organs provide complementary information about the same external event. For this reason, lipreading is used in understanding speech as it can compensate for interference from external noise and may resolve internal ambiguities of the auditory speech signal. A second reason is that there is internal 'drift' or 'error' within the individual senses that can be adjusted by cross-reference to other modalities. In the spatial domain of sensor-motor adaptation to optical-wedge prisms, this is already known for more than 100 years [30], but for speech, this kind of cross-reference to other modalities has been reported only very recently [3]. In both cases, though, there is a perceptual adjustment induced by a deviance between two information sources that the brain tries to reduce. The present study extends these findings by showing that this kind of adjustment not only occurs for auditory, but also for visual speech.

Our findings are also of relevance for the neural mechanisms involved in multisensory processing of audiovisual speech. Neuroimaging and electrophysiological studies have found audiovisual interactions in multimodal areas such as the superior temporal sulcus (STS) and sensory-specific areas including the auditory and visual cortices [4,6]. It has been proposed that the unimodal inputs are initially integrated in STS and that interactions in the primary auditory and visual cortices reflect feedback from STS [7]. On this account, interactions in the primary cortex are presumably mediated by the STS via backward projections [4]. Besides STS, motor regions of planning and execution (Broca's area, premotor cortex, and anterior insula) could be involved via the so-called mirror neurons (e.g. [12,15,18,25]). Broca's area is proposed to be a homologue of the macaque inferior premotor cortex (area F5) where mirror neurons are situated that discharge upon action and perception of goal-directed hand or mouth movements. The presumed function of these mirror neurons is to mediate imitation and aid action and understanding [21]. Broca's area is not only involved in the production of speech, but is also activated during silent lipreading [9] and passive listening to speech [37]. On this view, activation of mirror neurons in Broca's area may facilitate a link between auditory and visual speech inputs and the corresponding motor representations. In line with this notion, it has been reported that recalibration of auditory 'sine-wave speech' by lipread information occurs only if the sine-wave tokens were perceived as speech, but not if they were perceived as non-speech sounds [31] most likely because in the latter case, there was no link to articulatory motor programmes. Vision may thus affect auditory processing via articulatory motor programs of the observed speech acts [5], and as demonstrated here, it is conceivable that this effect is bi-directional in nature.

References

[1] S. Anstis, Motion perception in the frontal plane: sensory aspects, in: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, vol. 1, Wiley, New York, 1986 (Chapter 16).

[2] P. Bertelson, Ventriloquism: a case of cross-modal grouping, in: G. Aschersleben, T. Bachmann, J. Müsseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events, Elsevier, Amsterdam, 1999, pp. 347–362.

[3] P. Bertelson, J. Vroomen, B. De Gelder, Visual recalibration of auditory speech identification: a McGurk aftereffect, Psychol. Sci. 14 (2003) 592–597.

[4] J. Besle, A. Fort, C. Delpuech, M.H. Giard, Bimodal speech: early suppressive visual effects in human auditory cortex, Eur. J. Neurosci. 20 (2004) 2225–2234.

[5] D.E. Callan, J.A. Jones, K. Munhall, A.M. Callan, C. Kroos, E. Vatikiotis-Bateson, Neural processes underlying perceptual enhancement by visual speech gestures, Neuroreport 14 (2003) 2213–2218.

[6] D.E. Callan, J.A. Jones, K. Munhall, C. Kroos, A.M. Callan, E. Vatikiotis-Bateson, Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information, J. Cogn. Neurosci. 16 (2004) 805–816.

[7] G.A. Calvert, M.J. Brammer, E.T. Bullmore, R. Campbell, S.D. Iversen, A.S. David, Response amplification in sensory-specific cortices during crossmodal binding, Neuroreport 10 (1999) 2619–2623.

[8] G.A. Calvert, E.T. Bullmore, M.J. Brammer, R. Campbell, S.C. Williams, P.K. McGuire, P.W. Woodruff, S.D. Iversen, A.S. David, Activation of auditory cortex during silent lipreading, Science 276 (1997) 593–596.

[9] R. Campbell, M. MacSweeney, S. Surguladze, G. Calvert, P. McGuire, J. Suckling, M.J. Brammer, A.S. David, Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning), Brain Res. Cogn. Brain Res. 12 (2001) 233–243.

[10] C. Colin, M. Radeau, A. Soquet, D. Demolin, F. Colin, P. Deltenre, Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory, Clin. Neurophysiol. 113 (2002) 495–506.

[11] M.O. Ernst, H.H. Bülthoff, Merging the senses into a robust percept, Trends Cogn. Sci. 8 (2004) 162–169.

[12] M.H. Giard, F. Peronnet, Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study, J. Cogn. Neurosci. 11 (1999) 473–490.

[13] J.J. Gibson, Adaptation, after-effects and contrast in the perception of curved lines, J. Exp. Psychol. 18 (1933) 1–31.

[14] S. Hidaka, Y. Manaka, W. Teramoto, Y. Sugita, R. Miyauchi, J. Gyoba, Y. Suzuki, Y. Yukio Iwaya, Alternation of sound location induces visual motion perception of a static object, PLoS One 4 (2009) 1–6.

[15] V. Klucharev, R. Möttönen, M. Sams, Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception, Brain Res. Cogn. Brain Res. 18 (2003) 65–75.

[16] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.

[17] R. Möttönen, C.M. Krause, K. Tiippana, M. Sams, Processing of changes in visual speech in the human auditory cortex, Brain Res. Cogn. Brain Res. 13 (2002) 417–425.

[18] V. Ojanen, R. Möttönen, J. Pekkola, I.P. Jääskeläinen, R. Joensuu, T. Autti, M. Sams, Processing of audiovisual speech in Broca's area, Neuroimage 25 (2005) 333–338.

[19] J. Pekkola, V. Ojanen, T. Autti, I.P. Jaaskelainen, R. Mottonen, A. Tarkiainen, M. Sams, Primary auditory cortex activation by visual speech: an fMRI study at 3 T, Neuroreport 16 (2005) 125–128.

[20] M. Radeau, P. Bertelson, Auditory–visual interaction and the timing of inputs. Thomas (1941) revisited, Psychol Res. Psych. Fo. 49 (1987) 17–22.

[21] G. Rizzolatti, L. Craighero, The mirror-neuron system, Annu. Rev. Neurosci. 27 (2004) 169–192.

[22] M. Roberts, Q. Summerfield, Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory, Percept. Psychophys. 30 (1981) 309–314.

[23] H.M. Saldaña, L.D. Rosenblum, Selective adaptation in speech perception using a compelling audiovisual adaptor, J. Acoust. Soc. Am. 95 (1994) 3658–3661.

[24] M. Sams, R. Aulanko, M. Hämäläinen, R. Hari, O.V. Lounasmaa, S.T. Lu, J. Simola, Seeing speech: visual information from lip movements modifies activity in the human auditory cortex, Neurosci. Lett. 127 (1991) 141–145.

[25] J.I. Skipper, H.C. Nusbaum, S.L. Small, Listening to talking faces: motor cortical activation during speech perception, Neuroimage 25 (2005) 76–89.

[26] J.J. Stekelenburg, J. Vroomen, Neural correlates of multisensory integration of ecologically valid audiovisual events, J. Cogn. Neurosci. 19 (2007) 1964–1973.

[27] J. Tuomainen, T.S. Andersen, K. Tiippana, M. Sams, Audio–visual speech perception is special, Cognition 96 (2005) B13–22.

[28] S. van Linden, J. Vroomen, Recalibration of phonetic categories by lipread speech versus lexical information, J. Exp. Psychol. Human 33 (2007) 1483–1494.

[29] V. van Wassenhove, K.W. Grant, D. Poeppel, Visual speech speeds up the neural processing of auditory speech, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 1181–1186.

[30] H. von Helmholz, Treatise on Physiological Optics, Dover Publications, New York, 1867/1925, pp. 1867–1925.

[31] J. Vroomen, M. Baart, Phonetic recalibration only occurs in speech mode, Cognition 110 (2009) 254–259.

[32] J. Vroomen, M. Baart, Recalibration of phonetic categories by lipread speech: measuring aftereffects after a twenty-four hours delay, Lang. Speech 52 (2009) 341–350.

[33] J. Vroomen, J.J. Stekelenburg, Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli, J. Cogn. Neurosci., in press.

[34] J. Vroomen, S. van Linden, B. de Gelder, P. Bertelson, Visual recalibration and selective adaptation in auditory–visual speech perception: contrasting build-up courses, Neuropsychologia 45 (2007) 572–577.

[35] J. Vroomen, S. van Linden, M. Keetels, B. de Gelder, P. Bertelson, Selective adaptation and recalibration of auditory speech by lipread information: dissipation, Speech Commun. 44 (2004) 55–61.

[36] B.E. Walden, R.A. Prosek, A.A. Montgomery, C.K. Scherr, C.J. Jones, Effects of training on the visual recognition of consonants, J. Speech Hear. Res. 20 (1977) 130–145.

[37] S.M. Wilson, A.P. Saygin, M.I. Sereno, M. Iacoboni, Listening to speech activates motor areas involved in speech production, Nat. Neurosci. 7 (2004) 701–702.