

# Phonetic recalibration does not depend on working memory

Martijn Baart · Jean Vroomen

Received: 6 January 2010 / Accepted: 14 April 2010 / Published online: 1 May 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Listeners use lipread information to adjust the phonetic boundary between two speech categories (phonetic recalibration, Bertelson et al. 2003). Here, we examined phonetic recalibration while listeners were engaged in a visuospatial or verbal memory working memory task under different memory load conditions. Phonetic recalibration was—like selective speech adaptation—not affected by a concurrent verbal or visuospatial memory task. This result indicates that phonetic recalibration is a low-level process not critically depending on processes used in verbal- or visuospatial working memory.

**Keywords** Phonetic recalibration · Selective speech adaptation · Verbal working memory · Visuospatial working memory · Lipread aftereffects

In natural speech, there are other information sources besides the auditory signal that facilitate perception of the spoken message. For example, viewing a speaker's articulatory movements (i.e. lipreading) is known to improve auditory speech intelligibility (e.g. Erber 1974), especially when the auditory input is ambiguous (Sumby and Pollack 1954). More recent work has demonstrated that listeners also use lipread information to adjust the phonetic boundary between two speech categories (Bertelson et al. 2003; Vroomen et al. 2004, 2007; van Linden and Vroomen 2007, 2008; Vroomen and Baart 2009b). For example, listeners

exposed to an auditory ambiguous speech sound halfway between /b/ and /d/ (i.e. A? for auditory ambiguous) that is combined with the video of a speaker articulating either /b/ or /d/ (Vb and Vd for visual /b/ or /d/, respectively) report in a subsequently delivered auditory-only test more 'b'-responses after exposure to A?Vb than after A?Vd, as if they had learned to label the ambiguous sound in accordance with the lipread information (i.e., *phonetic recalibration*). Lipread-induced recalibration of phonetic categories has now been demonstrated many times (Vroomen et al. 2004, 2007; van Linden and Vroomen 2007, 2008; Vroomen and Baart 2009a, b) and has also been demonstrated to occur if the disambiguating information stems from lexical knowledge about the possible words in the language rather than from lipread information (e.g. Norris et al. 2003; Kraljic and Samuel 2005, 2006, 2007; van Linden and Vroomen 2007).

The mechanism underlying phonetic recalibration though is at present largely unknown. A recent functional magnetic resonance imaging (fMRI) study (Kilian-Hütten et al. 2008) using the same stimuli and design as in Bertelson et al. (2003) showed that the trial-by-trial variation in the amount of recalibration could be predicted from activation in the middle/inferior frontal gyrus (MFG/IFG) and the inferior parietal cortex. These brain areas are also known to be involved in verbal working memory (Jonides et al. 1998), and it might thus be conceivable that phonetic recalibration shares neural underpinnings with verbal working memory. Alternatively, though, there is behavioral and neurophysiological evidence which shows that lipreading has profound effects on speech perception at very early processing levels and that the effect is quite automatic (e.g. McGurk and MacDonald 1976; Massaro 1987, 1998; Colin et al. 2002; Möttönen et al. 2002; Soto-Faraco et al. 2004). On this view, it may seem more

M. Baart · J. Vroomen (✉)  
Department of Medical Psychology and Neuropsychology,  
Tilburg University, Warandelaan 2, P. O. Box 90153,  
5000 LE Tilburg, The Netherlands  
e-mail: j.vroomen@uvt.nl  
URL: <http://spitswww.uvt.nl/~vroomen>

likely that lipread-induced recalibration would not rely on high-level neural resources used for working memory, because it is basically a low-level process operating in an automatic fashion.

To examine whether phonetic recalibration and working memory indeed share common resources, we measured phonetic recalibration while participants were engaged in a working memory task. In the literature on working memory, a distinction is usually made between a verbal and a visuospatial component (e.g. Baddeley and Hitch 1974; Baddeley and Logie 1999), which rely on distinct neural structures. For example, Smith, Jonides and Koeppel (1996) showed primarily left-hemisphere activation during a verbal memory task, whereas the visuospatial task mainly activated right-hemisphere regions.

As a control for general disturbances caused by the dual task, we also examined whether the verbal and spatial memory task would interfere with selective speech adaptation. Selective speech adaptation, first demonstrated by Eimas and Corbit (1973), depends on the repeated presentation of a particular speech sound that causes a reduction in the frequency with which that token is reported in subsequent identification trials. Since its introduction, many questions have been raised about the nature underlying this effect. Originally, it was thought to reflect a fatigue of some hypothetical ‘linguistic feature detectors’, but others argued that it reflects a shift in criterion (e.g. Diehl et al. 1978), or a combination of both (Samuel 1986). Still others (e.g. Ganong 1978) showed that the size of selective speech adaptation depends upon the degree of spectral overlap between the adapter and test sound and that most of the effect is auditory rather than phonetic in nature. Moreover, selective speech adaptation is automatic as it is unaffected by a secondary online arithmetic or rhyming task (Samuel and Kat 1998). Following this line of reasoning, we did not expect our working memory task to interfere with selective speech adaptation.

To induce phonetic recalibration and selective speech adaptation, we used the same stimuli and procedures as in Bertelson et al. (2003). Participants were presented with multiple short blocks of eight audiovisual exposure trials immediately followed by six auditory-only test trials. During each exposure-test block, participants tried to memorize a set of previously presented letters for the verbal memory task or a motion path of a moving dot for the spatial task. The difficulty of the secondary memory task was increased across three groups of participants up until the point that performance on both memory tasks was about equal, sufficiently above chance level but below ceiling.

To the extent that phonetic recalibration shares mechanisms with working memory, one might expect more

interference from the verbal rather than spatial memory task because lipreading also relies primarily on activation in the left hemisphere (Calvert and Campbell 2003). Moreover, interference should increase if the memory task becomes more demanding. Alternatively, though, if recalibration is, like selective speech adaptation, a low-level process running in an automatic fashion, then neither the verbal nor the spatial memory task should interfere with recalibration.

## Method

### Participants

Sixty-six native speakers of Dutch (mean age = 21 years) with normal hearing and normal/corrected to normal vision participated, twenty-two in each of three memory load conditions. All participants gave their written informed consent prior to testing, and the experiment was conducted according to the Declaration of Helsinki.

### Stimuli

#### *Adapters*

The audiovisual adapter stimuli are described in detail in Bertelson et al. (2003). In short, the audio track of audiovisual recordings of a male speaker of Dutch pronouncing /aba/ and /ada/ were synthesised into a nine-step /aba/-/ada/ continuum in equal Mel-steps. To induce recalibration, the token from the middle of the continuum (A?) was dubbed onto both videos so as to create A?Vb and A?Vd. To induce selective speech adaptation, two audiovisual congruent adapters were created by dubbing the continuum endpoints onto the corresponding videos for AbVb and AdVd. As test stimuli served the most ambiguous sound on the continuum /A?/ and its immediate continuum neighbors /A?-1/ (more ‘/aba/-like’) and /A?+1/ (more ‘/ada/-like’).

### Design and procedure

Participants were tested individually in a sound-attenuated and dimly lit booth. They sat at approximately 70 cm from a 17-inch CRT screen. The audio was delivered via two regular loudspeakers placed left and right of the monitor at 63 dBa (measured at ear level). The videos showed the speaker’s entire face from the throat up to the forehead and were presented against a black background in the center of the screen (W: 10.4 cm, H: 8.3 cm). Testing was spread out over two subsequent days. Half of the participants were tested for recalibration on the first day, and selective speech adaptation on the second day, for the other half of

the participants the order was reversed. On both days, participants were tested in three separate blocks. One was a single-task adaptation procedure that served as baseline, the others were dual-task procedures using a visuospatial or a verbal memory task. Block order was counterbalanced across participants in a Latin square.

#### Recalibration/selective adaptation procedure

To induce recalibration, participants were exposed to eight repetitions (ISI = 425 ms) of either A?Vb or A?Vd. The exposure phase was immediately followed by an auditory-only test containing the ambiguous test stimulus /A?/, and its immediate neighbors on the continuum /A?-1/ and /A?+1/. These three test stimuli were presented twice in random order. After each test trial, participants had to indicate whether they heard /aba/ or /ada/ by pressing the corresponding 'b'- or 'd'-key on a response box. The next test trial was delivered 1,000 ms after a key press. There were sixteen exposure-test blocks (eight for A?Vb, and eight for A?Vd), all delivered in pseudo-random order.

The procedure to induce selective speech adaptation was exactly the same as for recalibration, except that participants were exposed to AbVb and AdVd. To ensure that participants attended the lipread videos during exposure, they were instructed—as in previous studies—to indicate whether they noticed an occasional small white dot on the upper lip of the speaker (12 px in size, 120 ms in duration).

#### Working memory tasks

In an attempt to equate task difficulty of the verbal and visuospatial memory tasks, we had to manipulate the set size of the memory items in a non-symmetrical way.

Verbal items were easier to remember than the visuospatial ones and for this reason, the number of memory items in both tasks differed as specified below.

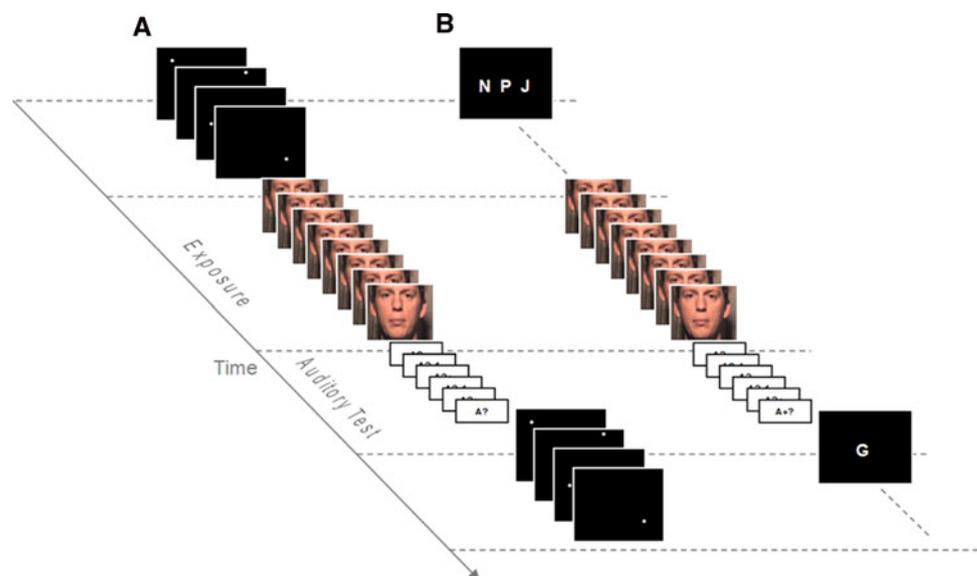
#### The visuospatial task

For the visuospatial task, each exposure-test block was preceded by a newly generated random path of a white dot ( $\emptyset = .4$  cm) that moved across a dark screen in three (for the low-memory load group) or four (for the intermediate- and high-memory load groups) steps. Each dot was presented for 500 ms. Participants were instructed to carefully attend to the target path and to remember it by covert repetition throughout the entire exposure-test block that would follow the target path. The exposure-test block was delivered to induce and measure recalibration or selective speech adaptation 1,300 ms after the last dot had disappeared. Immediately after this exposure-test block, participants were then presented a spatial probe for which they indicated whether its motion path was the same or different as the target by pressing a 'yes'- or 'no'-key (see Fig. 1a). In half of the trials, the target and the probe were the same, in the other half of the trials, the probe differed by one dot.

#### The verbal memory task

For the verbal memory task, participants had to remember a string of three (the low-memory load group), five (the intermediate-memory load group) or seven (the high-memory load group) letters that appeared simultaneously in the center of the screen for 2,000 ms. Participants were instructed to covertly repeat the string of letters throughout the exposure-test block that would follow. After the exposure-test block, a one-letter test probe was presented

**Fig. 1** Schematic overview of an exposure-test block in the low-load memory condition. In the visuospatial memory task (a), the motion path of a dot had to be remembered during the audiovisual exposure—auditory-only test phase. The memory probe immediately followed the final test token. In the verbal task (b), three letters had to be remembered



**Table 1** Proportion of correctly recognized probes in the verbal and visuospatial memory task at low-, medium-, and high-memory loads

Memory task	% of correct probes		
	Low	Load medium	High
Visuospatial	86	78	82
Verbal	98	92	83

for which participants indicated whether it was one of the targets by pressing the ‘yes’- or ‘no’-key (Fig. 1b). Half of the trials required a ‘yes’-response. The target letters were chosen from 16 consonants of the Latin alphabet, excluding ‘B’ and ‘D’, because they made up the crucial phonetic contrast. All letters were displayed in capitals (font type: Arial; size: 1.3(W) by 1.6(H) cm; spacing: 2.0 cm).

## Results

### Performance on the memory tasks

The average number of correct responses in the verbal and spatial memory task under the three load conditions is presented in Table 1. In the ANOVA on the percentage of correct responses, the main effect of task,  $F(1,64) = 40.40$ ,  $P < .001$ , showed that verbal probes were recognized somewhat better than the spatial probes, (91 vs. 82%, respectively, with chance level at 50%). There was also a main effect of load,  $F(1,64) = 23.30$ ,  $P < .001$ , because recognition became worse when load increased. There was an interaction between memory load and task;  $F(1,64) = 15.24$ ,  $P < .001$ , as increasing the memory load had a bigger impact on the verbal task (where set size was increased from 3 to 7 items) than the spatial task (where the target path was increased from 3 to 4 steps from low to medium, and remained at 4 during high load). As intended, in the high-load condition overall performance for the verbal and spatial task were not different from each other ( $P = .88$ ), so task difficulty was equated here. The results for the memory task confirm that participants were indeed paying attention to the task as performance was well above chance. Moreover, increasing memory load made the task more difficult, so it was not too easy. This pattern therefore provides a platform to answer the main question, namely whether increasing memory load interferes with phonetic recalibration.

### Performance on speech identification

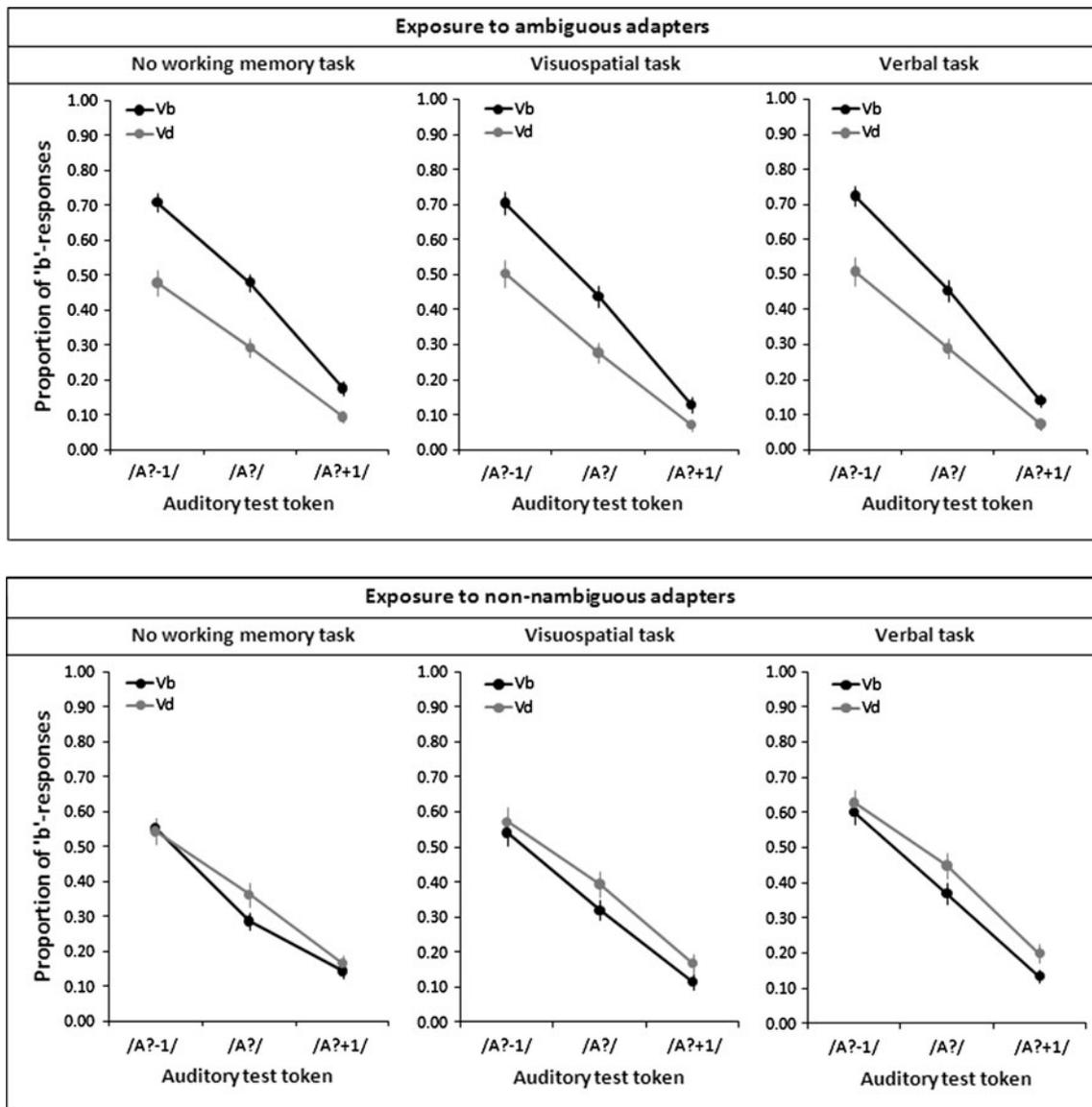
The data of the speech identification trials were analyzed as in previous studies by computing aftereffects (Bertelson et al. 2003; Vroomen and Baart 2009a). First, the average

number of ‘b’-responses as a function of the test token was calculated for each participant. The group-averaged data are presented in Fig. 2. The data in this figure are averaged across the three memory load groups because preliminary analyses showed that memory load did not affect performance in any rational way (all  $F$ 's with load as factor  $< 1$ ). As is clearly visible, there were more ‘b’-responses for the ‘b-like’ A?–1 token than the more ‘d-like’ A?+1 token. More interestingly, there were *more* ‘b’-responses after exposure to A?Vb than A?Vd (indicative of recalibration), whereas there were *fewer* b-responses after exposure to AbVb than AdVd (indicative of selective speech adaptation), thus replicating the basic results for recalibration and selective speech adaptation reported before.

To quantify these aftereffects, the proportion of ‘b’-responses following exposure to Vd was subtracted from exposure to Vb, thereby pooling over test tokens. Recalibration (A?Vb–A?Vd) manifested itself as *more* ‘b’-responses following exposure to A?Vb than A?Vd; whereas for selective speech adaptation (AbVb–AdVd), there were fewer ‘b’- responses after exposure to AbVb than AdVd (see Table 2). Most importantly, none of these aftereffects was modulated by either of the two secondary memory tasks. This was tested in a 2 (adapter sound: ambiguous/non-ambiguous)  $\times$  3 (task: no/visuospatial/verbal)  $\times$  3 (memory load: low/medium/high) ANOVA on the aftereffects with memory load as a between-subjects variable, and adapter sound and task as within-subjects variables. There was a main effect of adapter sound because exposure to the ambiguous adapter sounds induced positive aftereffects (recalibration), whereas exposure to the non-ambiguous sounds induced negative aftereffects (selective speech adaptation),  $F(1,64) = 27.33$ ,  $P < .001$ . Crucially, there was no effect of task;  $F(2,128) < 1$ , memory load;  $F(1,64) < 1$ , nor was there a higher order interaction between any of these variables (all  $P$ 's were at least  $> .3$ ). Aftereffects indicative of recalibration and selective speech adaptation were thus *unaffected* by whether participants were trying to remember letters or a visuospatial path during the exposure and test phase.

## Discussion

The present study indicates that a concurrent working memory task does not interfere with lipread-induced phonetic recalibration. Participants readily adapted their interpretation of an initially ambiguous sound based on lipread information, but this occurred independent of whether they were engaged in a demanding verbal or spatial working memory task. This suggests that phonetic recalibration is—like selective speech adaptation (Samuel and Kat 1998)—a low-level process that occurs in an



**Fig. 2** Proportion of ‘b’-responses after exposure to A?Vb and A?Vd (upper panels) and AbVb and AdVd (lower panels) for the single and dual tasks. Data are averaged over memory load. Error bars represent one standard error of the mean

**Table 2** Aftereffects after exposure to ambiguous and non-ambiguous adapter sounds while remembering verbal or spatial items at three loads

Memory task	Ambiguous adapter sound			Non-ambiguous adapter sound		
	Low	Load medium	High	Low	Load medium	High
No task	.15	.18	.16	-.04	-.04	-.02
Visuospatial	.15	.14	.12	-.08	-.05	-.02
Verbal	.14	.11	.17	-.07	-.06	-.05

automatic fashion. This finding is in line with other research that demonstrates that the online integration of auditory and visual speech is automatic (McGurk and MacDonald 1976; Massaro 1987; Campbell et al. 2001; Näätänen 2001; Colin et al. 2002; Möttönen et al. 2002;

Calvert and Campbell 2003; Besle et al. 2004; Callan et al. 2004; Soto-Faraco et al. 2004).

As a counterargument, it might be argued that the memory tasks were simply too easy to affect phonetic recalibration and selective speech adaptation. Against this

interpretation, though, is that increasing the memory load of the concurrent task *did* affect probe recognition. In the highest load conditions of the spatial and verbal memory task, recognition rate was at  $\sim 82\%$ , which is well above chance level, but far from being perfect. Participants were thus likely engaged in the memory task, yet it had no effect on phonetic recalibration or selective speech adaptation.

Yet, another counterargument is that one cannot be sure that participants were actively engaged in covertly repeating the memory items while they were exposed to the audiovisual speech tokens that supposedly drive recalibration. Admittedly, the critical part of the exposure phase that induces recalibration—the part in which a participant hears an ambiguous segment while seeing another phonetic segment—is very short, and there is no guarantee that participants were—at that specific time—actually engaged in repeating the memory items. Unfortunately, we cannot offer an obvious solution for this because it is a very general problem in dual-task paradigms where there is always uncertainty about strategic effects in performing the primary and secondary task. One might, as an alternative, have used a more demanding online task that allows one to keep track of performance during the exposure phase. Participants might for example track a concurrent visual stimulus while being exposed to the lipread information, as eye-tracking is relatively easy to measure (see e.g. Alsius et al. 2005). However, a disadvantage of this method is that the visual tracking task as such may interfere with lipreading, so there is interference at the sensory level rather than at the level at which phonetic recalibration occurs. Participants might thus simply not see the critical lipread information when simultaneously engaged in a visual tracking task. Other studies on audiovisual speech using this dual task have indeed found that an additional visual task (tracking a moving leaf over a speaking face) can interfere with lipreading (e.g. Tiippana et al. 2004), thus preventing any firm conclusion about whether attention affects cross-modal information integration rather than lipreading itself. A recent report on spatial attention (i.e. attending one out of two faces presented on the left and right of fixation) also confirms that endogenous attention affects lipreading rather than multisensory integration (Andersen et al. 2009).

Alternatively, one could also use a secondary task that does not interfere with the auditory and visual sensory requirements of the primary task, like, for instance, a tactile task. In a study by Alsius et al. (2007), it was indeed reported that the percentage of illusory McGurk responses decreased when participants were concurrently performing a difficult tactile task (deciding whether two taps were finger-symmetrical with the preceding trial). As already

argued, this result by itself does not unequivocally imply that the tactile secondary task had an effect on audiovisual integration per se, because the task may also interfere with unimodal processing of the lipread information, thus before audiovisual integration did take place. However, Alsius et al. (2005) and (2007), included auditory-only and visual-only baseline conditions in which participants repeated the word they had just heard or lipread. The authors did not find a difference in the unimodal baseline conditions between the single and dual tasks, which made them refute the idea that the secondary task affected lipreading rather than audiovisual integration. Here, we acknowledge that it remains for future research to examine whether a concurrent tactile task would also affect lipread-induced phonetic recalibration.

From a broader perspective, there is a current debate in the literature about the extent to which intersensory integration requires attentional resources. Some have argued that intersensory integration depends on attentional resources (e.g. Alsius et al. 2005; Fairhall and Macaluso 2009; Talsma et al. 2007), while others have argued it does not (e.g. Bertelson et al. 2000; Massaro 1987; Soto-Faraco et al. 2004; Vroomen et al. 2001a, b). Admittedly, the current experiment did not measure the role of attention as such, but being simultaneously engaged in two tasks is usually taken to imply that available attentional resources were divided across the two tasks. Given that there was no effect of the secondary task on lipread-induced recalibration, it appears that the present findings fit better within the perspective that multisensory integration is unconstrained by attentional resources. This finding also fits well with the observation that a face displaying an emotion has profound effects on auditory emotion-labeling but yet again, this effect occurs independent of whether or not listeners were instructed to add numbers, count the occurrence of a target digit in a rapid serial visual presentation or were asked to judge the pitch of a tone as high or low (Vroomen et al. 2001b). Similarly, in the spatial domain it has been demonstrated that vision can bias sound localization (i.e. the ventriloquist effect, e.g. Radeau and Bertelson 1974; Bertelson 1999), but this cross-modal bias occurs irrespective of where endogenous (Bertelson et al. 2000) or exogenous spatial attention is directed (Vroomen et al. 2001a).

To conclude, the data demonstrate that during lipread-induced phonetic recalibration, the auditory and visual signals were integrated into a fused percept that left longer-lasting traces. Apparently, listeners learned to interpret an initially ambiguous sound because there was lipread information that was used to disambiguate that sound. This phenomenon is—like selective speech adaptation—likely a low-level phenomenon that does not seem to depend on processes used in spatial or verbal working memory tasks.

We acknowledge, though, that at this point, the dual-task method leaves more than one interpretation open, and it appears that there is no other solution than running more experiments with different tasks.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol* 15:839–843
- Alsius A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. *Exp Brain Res* 183:399–404
- Andersen TS, Tiippana K, Laarni J, Kojo I, Sams M (2009) The role of visual attention in audiovisual speech perception. *Speech Commun* 51:184–193
- Baddeley AD, Hitch G (1974) Working memory. In: Bower GH (ed) *The psychology of learning and motivation: Advances in research and theory*, vol 8. Academic Press, New-York, pp 47–89
- Baddeley AD, Logie RH (1999) Working memory: the multiple-component model. In: Miyake A, Shah P (eds) *Models of working memory: mechanisms of active maintenance and executive control*. Cambridge University Press, New York, pp 28–61
- Bertelson P (1999) Ventriloquism: A case of cross-modal grouping. In: Aschersleben G, Bachmann T, Müsseler J (eds) *Cognitive contributions to the perception of spatial and temporal events*. Elsevier, Amsterdam, pp 347–362
- Bertelson P, Vroomen J, de Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys* 62:321–332
- Bertelson P, Vroomen J, De Gelder B (2003) Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci* 14:592–597
- Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20:2225–2234
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cognitive Neurosci* 16:805–816
- Calvert GA, Campbell R (2003) Reading speech from still and moving faces: the neural substrates of visible speech. *J Cognitive Neurosci* 15:57–70
- Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001) 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* 12:233–243
- Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P (2002) Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clin Neurophysiol* 113:495–506
- Diehl RL, Elman JL, McCusker SB (1978) Contrast effects on stop consonant identification. *J Exp Psychol Human* 4:599–609
- Eimas PD, Corbit JD (1973) Selective adaptation of linguistic feature detectors. *Cognitive Psychol* 4:99–109
- Erber NP (1974) Auditory-visual perception of speech: A survey. In: Nielsen HB, Kampp E (eds) *Visual and audio-visual perception of speech*. Almquist & Wiksell, Stockholm, Sweden
- Fairhall SL, Macaluso E (2009) Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci* 29:1247–1257
- Ganong WF (1978) The selective adaptation effects of burst-cued stops. *Percept Psychophys* 24:71–83
- Jonides J, Schumacher EH, Smith EE, Koeppel RA, Awh E, Reuter-Lorenz PA, Marhuetz C, Willis CR (1998) The role of parietal cortex in verbal working memory. *J Neurosci* 18:5026–5034
- Kilian-Hütten NJ, Vroomen J, Formisano E (2008) One sound, two percepts: predicting future speech perception from brain activation during audiovisual exposure. *Neuroimage* 41, Supplement 1: S112
- Kraljic T, Samuel AG (2005) Perceptual learning for speech: is there a return to normal? *Cognitive Psychol* 51:141–178
- Kraljic T, Samuel AG (2006) Generalization in perceptual learning for speech. *Psychon B Rev* 13:262–268
- Kraljic T, Samuel AG (2007) Perceptual adjustments to multiple speakers. *J Mem Lang* 56:1–15
- Massaro DW (1987) *Speech perception by ear and eye: A paradigm for psychological inquiry*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ
- Massaro DW (1998) *Perceiving talking faces: from speech perception to a behavioral principle*. The MIT Press, Cambridge
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748
- Möttönen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res* 13:417–425
- Näätänen R (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent. *Psychophysiology* 38:1–21
- Norris D, McQueen JM, Cutler A (2003) Perceptual learning in speech. *Cognitive Psychol* 47:204–238
- Radeau M, Bertelson P (1974) The after-effects of ventriloquism. *Q J Exp Psychol* 26:63–71
- Samuel AG (1986) Red herring detectors and speech perception: in defense of selective adaptation. *Cognitive Psychol* 18:452–499
- Samuel AG, Kat D (1998) Adaptation is automatic. *Percept Psychophys* 60:503–510
- Smith EE, Jonides J, Koeppel RA (1996) Dissociating verbal and spatial working memory using PET. *Cereb Cortex* 6:11–20
- Soto-Faraco S, Navarra J, Alsius A (2004) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition* 92:B13–B23
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:679–690
- Tiippana K, Andersen TS, Sams M (2004) Visual attention modulates audiovisual speech perception. *Eur J Cogn Psychol* 16:457–472
- van Linden S, Vroomen J (2007) Recalibration of phonetic categories by lipread speech versus lexical information. *J Exp Psychol Human* 33:1483–1494
- van Linden S, Vroomen J (2008) Audiovisual speech recalibration in children. *J Child Lang* 35:809–822
- Vroomen J, Baart M (2009a) Phonetic recalibration only occurs in speech mode. *Cognition* 110:254–259
- Vroomen J, Baart M (2009b) Recalibration of phonetic categories by lipread speech: measuring aftereffects after a twenty-four hours delay. *Lang Speech* 52:341–350

- Vroomen J, Bertelson P, de Gelder B (2001a) The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept Psychophys* 63:651–659
- Vroomen J, Driver J, de Gelder B (2001b) Is cross-modal integration of emotional expressions independent of attentional resources? *Cognit Affect Behav Neurosci* 1:382–387
- Vroomen J, van Linden S, Keetels M, de Gelder B, Bertelson P (2004) Selective adaptation and recalibration of auditory speech by lipread information: dissipation. *Speech Commun* 44:55–61
- Vroomen J, van Linden S, de Gelder B, Bertelson P (2007) Visual recalibration and selective adaptation in auditory-visual speech perception: contrasting build-up courses. *Neuropsychologia* 45:572–577