ELSEVIER

Neuropsychologia 45 (2007) 572–577

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses

Jean Vroomen [a], Sabine van Linden [a], Béatrice de Gelder [a], Paul Bertelson [a,b,*]

[a] *Department of Psychology, Tilburg University, The Netherlands*

[b] *Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles, 50 Av. F. D. Roosevelt, CP 191, 1050 Brussels, Belgium*

Received 5 September 2005; received in revised form 8 December 2005; accepted 30 January 2006

Available online 10 March 2006

Abstract

Exposure to incongruent auditory and visual speech produces both *visual recalibration* and *selective adaptation* of auditory speech identification. In an earlier study, exposure to an ambiguous auditory utterance (intermediate between /aba/ and /ada/) dubbed onto the video of a face articulating either /aba/ or /ada/, recalibrated the perceived identity of auditory targets in the direction of the visual component, while exposure to congruent non-ambiguous /aba/ or /ada/ pairs created selective adaptation, i.e. a shift of perceived identity in the opposite direction [Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science*, *14*, 592–597]. Here, we examined the build-up course of the after-effects produced by the same two types of bimodal adapters, over a 1–256 range of presentations. The (negative) after-effects of non-ambiguous congruent adapters increased monotonically across that range, while those of ambiguous incongruent adapters followed a curvilinear course, going up and then down with increasing exposure. This pattern is discussed in terms of an asynchronous interaction between recalibration and selective adaptation processes.

Keywords: Auditory–visual speech; Speechreading; After-effect; Recalibration; Perceptual learning; Selective adaptation; McGurk effect

The question of how sensory modalities cooperate in forming a coherent representation of the environment is the focus of much current research. The major part of that work is carried out with conflict situations, in which incongruent information about potentially the same distal event is presented to different modalities (see reviews by Bertelson & de Gelder, 2004; De Gelder & Bertelson, 2003).

Exposure to such conflicting inputs produces two main effects: immediate biases and after-effects. By immediate biases are meant effects of incongruent inputs in a distracting modality on the perception of corresponding inputs in a target modality. For example, in the so-called ventriloquist illusion, the perceived location of target sounds is displaced toward light flashes delivered simultaneously at some distance, in spite of instructions to ignore the latter (Bertelson, 1999). After-effects (henceforth "AEs") are shifts in perception observed following exposure to an inter-modal conflict, when data in one or in both modalities are later presented alone. For the ventriloquism situation, unimodal sound localization responses are, after exposure to synchronized but spatially discordant sound bursts and light flashes, shifted in the direction of the distracting flashes (Radeau & Bertelson, 1974; Recanzone, 1998). The occurrence of AEs has generally been taken as implying that exposure to incongruence between corresponding inputs in different modalities *recalibrates* processing in one or both modalities in a way that eliminates (or at least reduces) the perceived discordance. Although immediate biases and recalibration have consistently been demonstrated for spatial conflict situations, the evidence has long been less complete for conflicts regarding event identities. Here, biases were often reported, but, for some time, no recalibration. The main example is the conflict resulting from the acoustic delivery of a particular speech utterance in synchrony with the optical presentation of a face articulating a visually incongruent utterance. As originally reported by McGurk and MacDonald (1976), this kind of situation generally produces strong immediate biases of the auditory percept towards the speechread distracter, a phenomenon now generally called "the McGurk effect". For instance, auditory /ba/ combined with

* Corresponding author. Tel.: +32 2 772 85 81; fax: +32 2 650 22 09.
E-mail address: pbrtlsn@ulb.ac.be (P. Bertelson).

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.01.031

visual /ga/ is often heard as /da/. On the other hand, no demonstration of AEs consequent upon exposure to McGurk situations had until recently been reported, and results in the literature (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994) were taken as implying that such exposure produces no recalibration, possibly revealing a basic difference between identity and spatial conflicts (Rosenblum, 1994).

Using a new type of adapting situation, we have however now succeeded in demonstrating the latter kind of recalibration (Bertelson, Vroomen, & de Gelder, 2003). Our exposure situation involved bimodal stimulus pairs in which the auditory component was each participant's most ambiguous speech utterance from an /aba/–/ada/ continuum (A?), and the visual component featured the articulation of either of the two end points, /aba/ or /ada/. Following the habitual conflict adaptation paradigm, auditory identification tests, using the ambiguous utterance and two slightly less ambiguous ones as material, were administered after exposure to bimodal adapters with either the /aba/ or the /ada/ visual component. As expected, /aba/ responses were more frequent after exposure with visual /aba/ than with visual /ada/, thus revealing recalibration.

Our reason for using an ambiguous auditory adapter was to avoid the occurrence of the so-called *selective speech adaptation* phenomenon, in which repeated exposure to a non-ambiguous auditory speech utterance causes a reduction in the frequency with which that utterance is reported on subsequent identification trials (Eimas & Corbit, 1973; Samuel, 1986). Selective speech adaptation is thus, like recalibration, an adaptation phenomenon that manifests itself by AEs but, unlike recalibration, does not depend on the co-occurrence of conflicting inputs in another modality. If our bimodal exposure had been run with unambiguous auditory utterances, e.g. auditory /aba/ paired with visual /ada/, the same outcome on post-test, more /ada/ responses, could have been equally attributed to selective speech adaptation from auditory /aba/ as to recalibration of speech identification by the visual distracter /ada/.

That exposure to bimodal pairs with unambiguous auditory speech utterances from our material can actually produce selective speech adaptation was demonstrated in the same study (Bertelson et al., 2003, Exp. 2) by exposing participants to congruent and unambiguous audio-visual pairs, either auditory /aba/ combined with visual /aba/, or auditory /ada/ combined with visual /ada/. In this new condition, exposure effectively resulted in a reduction of the proportion of responses consistent with the bimodal adapter. Fewer /aba/ responses occurred after exposure to bimodal /aba/ than to bimodal /ada/, the outcome opposite the one obtained when the same visual /aba/ was paired with the ambiguous auditory utterance. The congruent visual component presumably played no role in the causation of selective adaptation, but its presence made each congruent non-ambiguous adapting pair undistinguishable from the pair with the same visual component and the ambiguous auditory component, as was shown in separate identification tests.

Additional evidence for the dissociation between two adaptation phenomena was provided more recently in a study showing that they *dissipate* following different courses (Vroomen, van Linden, Keetels, de Gelder, & Bertelson, 2004). The present study is focused on the *build-up* of the AEs across successive presentations of the bimodal adapters of our original study (Bertelson et al., 2003). Two of these, making up the *ambiguous sound condition*, consisted of the participant's most ambiguous auditory utterance A?, paired across successive presentations either with visual /aba/ (pair A?Vb) or with visual /ada/ (pair A?Vd). The other two adapters, making up the *non-ambiguous sounds condition*, consisted of auditory /aba/ paired with visual /aba/ (pair AbVb) and of auditory /ada/ paired with visual /ada/ (pair AdVd). Following the earlier findings, the ambiguous sound condition was expected to produce no selective speech adaptation, because of the ambiguity of the auditory component, but to cause recalibration in the direction of the incongruent visual component. In contrast, the non-ambiguous sounds condition was expected to produce selective adaptation, because of the non-ambiguous quality of each auditory component, but no recalibration, because of the absence of phonetic incongruence between auditory and visual components. The adapters were presented in continuous series of trials, and auditory AEs were measured at several successive points during each series. A first group of participants was tested with adaptation blocks running to 64 trials. Their results revealed an unexpected reversal in the build-up course of adaptation in the ambiguous sound condition. To check on this finding, the number of exposure trials was extended to 256 for a second group of participants.

1. Methods

1.1. Materials

Details of the stimuli have been provided in an earlier paper (Vroomen et al., 2004). In short, a 9-point /aba/–/ada/ speech continuum was created by varying the frequency of the second (F2) formant in equal steps. The end-point auditory utterances and the individually determined most ambiguous one were dubbed onto the video of a face that articulated /aba/ or /ada/.

1.2. Participants

Two groups of 25 students from Tilburg University participated in one experimental session. Those in Group 64 were administered 64 trials long exposure blocks, and those in Group 256, 256 trials long blocks.

1.3. Procedure

For both groups, the session involved three successive phases: *calibration*, then *pre-tests* followed by a bimodal *audio-visual exposure* phase, interspersed with *post-test trials*.

The *calibration phase* served to determine, for each participant individually, the sound on the continuum that was nearest to her/his /aba/–/ada/ phoneme boundary. It consisted of 98 trials in which each of the nine sounds was presented in random order at 1.5 s inter-trials intervals. Sounds from the middle of the continuum were presented more often than those from the extremes (6, 8, 14, 14, 14, 14, 14, 8 and 6 presentations for each of the nine items, respectively). The participant classified the sound as /aba/ or /ada/ by pressing one of two keys. The participant's 50% cross-over point was estimated via probit analysis, and the continuum item nearest to that point (A?) served as auditory component in the bimodal exposure trials of the ambiguous sound condition.

In the *pre-test phase*, the participant gave dichotomous key-pressing classification responses to her/his most ambiguous sound A?, as well as to its two immediate continuum neighbors (A? − 1 and A? + 1). These three test sounds were presented in balanced order across 20 successive triplets. The 60 presentations followed each other without interruption at 2.5 s ITIs.