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*

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.
© 2006 Elsevier Ltd. All rights reserved.

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

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.

The *audio-visual exposure* phase consisted of eight adaptation blocks, two for each of the four bimodal adapters AbVb, AdVd, A?Vb and A?Vd. For Group 64, the adapters were presented 64 times in each block at 1.5 s ITIs, and two triplets of auditory identification post-tests, identical to those in the pre-test phase, were interpolated after 1, 2, 4, 8, 16, 32, and 64 exposures. For Group 256, there were 256 exposures per block presented at 0.85 s ITIs, and post-tests were interpolated at the same locations as for Group 64, plus locations 128 and 256. No phonetic decisions were required during the audio-visual exposure phase, but participants had to press a special key on every presentation of a visual catch stimulus (a 12 pixels white spot flashed for 100 ms between the nose and the upper lip of the talker). Five such catch trials were interpolated at random moments during each block, in order to ensure attention to the face. Presentations of the different types of blocks were counterbalanced across participants.

2. Results

The individually determined most ambiguous auditory stimuli (A?) ranged between utterances 4 and 6. During bimodal exposure, participants detected 93% of the visual catch stimuli, indicating that they were effectively attending to the face. AEs were calculated (as in Bertelson et al., 2003), by the difference in the proportion of /aba/ responses obtained after exposure to respectively A?Vb and A?Vd (ambiguous sound condition) or after AbVb and AdVd (non-ambiguous sound condition). Recalibration thus manifests itself by positive AEs and selective adaptation by negative ones. Mean AEs as functions of number of exposure trials across each type of block are shown in Fig. 1.

As a first step, the data from each group were submitted to two separate two-factor (auditory ambiguity and number of exposures) MANOVAs. For Group 64, both main effects, auditory ambiguity, $F(1, 24) = 52.6$, $p < .001$, and number of exposures, $F(6, 144) = 6.37$, $p < .001$, and their interaction $F(6, 144) = 15.3$, $p < .001$, were significant. The results were identical for Group 256: condition, $F(1, 24) = 57.1$, $p < .001$, number of exposures, $F(8, 192) = 31.8$, $p < .001$, interaction, $F(8, 192) = 13.2$, $p < .001$. The effects of auditory ambiguity correspond to the fact that AEs were mainly positive with ambiguous adapters and mainly negative with non-ambiguous ones. The interactions reflect the fact,

clearly visible in Fig. 1, that AEs follow different courses in the two conditions, monotonically decreasing in the non-ambiguous sound condition and going up and then down in the ambiguous sound condition.

In the ambiguous sound condition, AEs appear to peak higher in Group 64 than in Group 256. To check on the significance of that difference, the data of Group 64 were entered together with those for the first 64 exposures of Group 256 into a two-factor (group and number of exposures) MANOVA. No significant main effect of group, $F(1, 48) < 1$, nor any interaction with number of exposures, $F(6, 288) = 1.52$, $p > .10$, emerged. Thus, the observed difference probably resulted from random variations among participants in the two groups. A similar absence of group difference could be expected for the non-ambiguous condition on the basis of the data in Fig. 1, and was confirmed by MANOVA, both $F_s < 1$.

Given the absence of significant difference, the data from the two groups could be pooled and the resulting values (shown in Fig. 2) submitted to two General Linear Model (GLM) analyses, allowing the examination of trends. For the ambiguous sound condition, the analysis produced a significant quadratic component, $F(1, 49) = 7.34$, $p < .01$, and the linear component, $F(1, 49) = 1.65$, $p > .20$, was non-significant. The quadratic component reflects the fact that the AE raised, reached a plateau and then went down. For the non-ambiguous condition (lower part of Fig. 2), GLM produced a highly significant linear component, $F(1, 49) = 91.2$, $p < .001$, as well as significant quadratic, $F(1, 49) = 21.6$, $p < .001$, and cubic, $F(1, 49) = 11.6$, $p < .001$, ones. The linear component reflects the monotonic decreasing slope of the curve and the two higher order components, its gradual flattening. Finally, application of GLM to the 64 to 256 exposures AEs of Group 256 produced significant linear trends ($p < .01$ for both conditions) and no higher order trends (all $F_s < 1$).
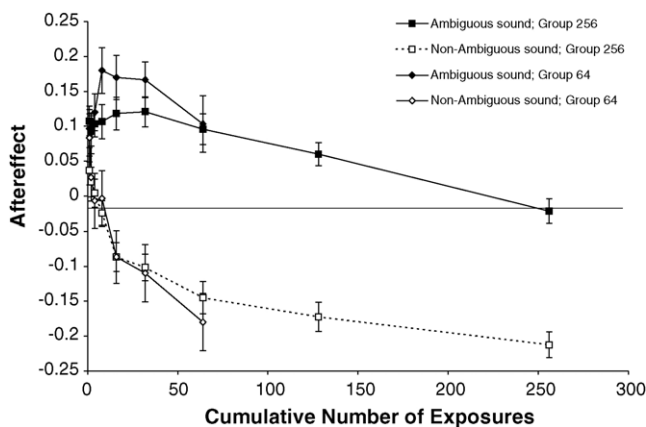

Fig. 1. Mean after-effects as functions of cumulative number of exposures in the ambiguous sound condition (adapters A?Vb and A?Vd) and the non-ambiguous sound condition (adapters AbVb and AdVd) for Group 64 (exposures 1–64) and Group 256 (exposures 1–256). Aftereffects are the differences between the proportions /aba/ responses obtained with adapters A?Vb and A?Vd (ambiguous condition) or with adapters AbVb and AdVd (non-ambiguous condition).
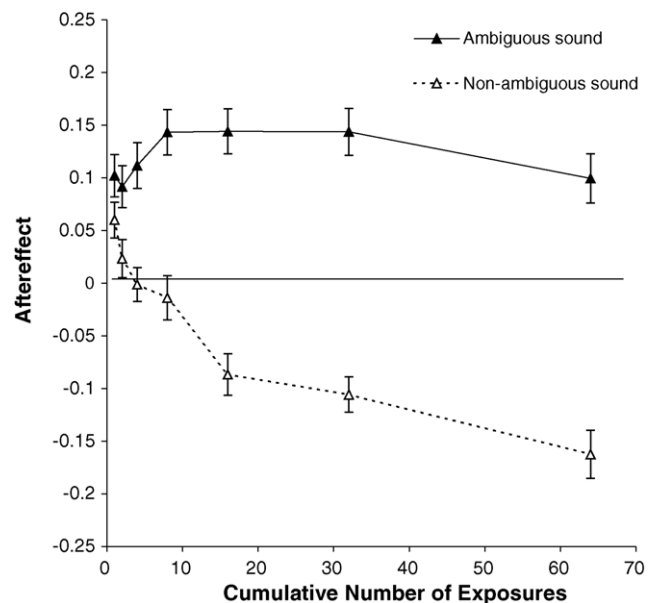

Fig. 2. Mean aftereffects as functions of cumulative number of exposures (1–64), for the pooled data of Groups 64 and 256, in the ambiguous sound condition (adapters A?Vb and A?Vd) and the non-ambiguous sound condition (adapters AbVb and AdVd).

Two somewhat unexpected aspects of the build-up courses deserve attention. Both concern the starting points of the curves. In the ambiguous sound condition, a substantial positive AE, significantly superior to zero, $t(49) = 5.98$, $p < .001$, already occurs after the single first presentation of the bimodal adapter. In the non-ambiguous sound condition, a significant positive AE, $t(49) = 2.98$, $p < .005$, occurs after the first presentation. It gives way to the expected negative values on succeeding exposures. Possible reasons for these effects will be examined in Section 3.

3. Discussion

The present experiment examined the way the contrasting auditory AEs obtained in our earlier studies (Bertelson et al., 2003; Vroomen et al., 2004), after exposure to bimodal pairs with respectively ambiguous and non-ambiguous auditory components, build-up. Two main results emerged.

First, the main directions in which AEs develop are the same as in the earlier experiments. After eight presentations (the level of exposure used throughout in the original study) AEs went in the direction of the visual distracter in the ambiguous sound conditions, and in the opposite direction (away from the congruent bimodal adapter) in the non-ambiguous sound conditions. This contrast, which was presented as demonstrating the dissociation between recalibration and selective speech adaptation, is thus replicated. The fact established in the original study (Bertelson et al., 2003) that in our material corresponding bimodal adapters differing only at the level of auditory ambiguity (like A?Vb and AbVb) are perceptually undistinguishable should be stressed again at this point. It carries the important implication that the contrasting AEs obtained in the two conditions cannot have originated in deliberate post-perceptual strategies, and must be of perceptual nature.

Second, the respective developments not only go in opposite directions, but also follow different courses, monotonically descending for non-ambiguous sounds, and curvilinear, with a rapid early build-up followed by a plateau and then a gradual decline, for ambiguous sounds.

In our initial paper (Bertelson et al., 2003) we proposed that the AEs obtained in the ambiguous sound condition reflected essentially recalibration, and those in the non-ambiguous sound condition, essentially selective adaptation. Let us now examine how the build-up results affect these proposals.

For the non-ambiguous sound condition, the monotonic descend of the curve is consistent with the interpretation in terms of cumulative selective adaptation, and the gradual deceleration of that descent suggests evolution toward some asymptotic value. The fact that the descending curve starts, after the first exposure trial, not at zero or already at some small negative level, but at a significantly positive one, may seem surprising. A possible explanation would be that presentation of a non-ambiguous (end-point) auditory utterance produces not only selective adaptation but also some priming or repetition effect, i.e. moving perception of the ambiguous test utterance in the direction of the just presented non-ambiguous one, the direction opposite that of selective adaptation. If the priming effect, in contrast to the cumulative selective adaptation, was constant from trial to

trial, it might overrun the latter on early presentations but be overtaken by it later on, thus producing the pattern observed in the figure.

For the ambiguous sound condition, the main finding is the curvilinear development course. That an initial positive growth gradually gives way to a decline is supported by the quadratic trend obtained for the pooled data over exposures 1–64, and the reality of the final decline receives additional support from the descending linear trend obtained over the last three post-tests of Group 256. What mechanism could produce such a pattern?

The ascending part of the curve in all probability reflects increasing recalibration. A question similar to the one discussed for the non-ambiguous sound data may be raised concerning the significant AE already present after the first exposure. The two cases are however not identical. In the non-ambiguous condition, the first trial AE went in the direction opposite the later selective adaptation, thus requiring a different explanation, like the one through a priming effect that we proposed. For the ambiguous condition, the first trial AE goes in the same positive direction as the later build-up, so that it can just be the effect of a very rapid, or one-trial, recalibration process. That priming would also play some role cannot be excluded on the basis of the present data, and the possibility should be a matter for future investigations.

Regarding the later decline, there is of course no apparent reason why a learning phenomenon like recalibration would *reverse* itself at some point. Some separate process must be involved here. The most likely possibility is a selective adaptation process running in parallel with recalibration and eventually counterbalancing it. This process could start as soon as some sufficient exposure to non-ambiguous sounding inputs has occurred. A basis for selective adaptation can be provided on each trial since pairing the ambiguous utterance with a (non-ambiguous) visual component makes it sound non-ambiguous (through the McGurk effect). Of course, the same bimodal pair also produces recalibration because of the discrepancy between the ambiguous utterance and the non-ambiguous visual component. Whether the progressive recalibration then makes an additional contribution to the accumulation of selective adaptation is a possibility, but one on which our data provide no evidence.

In conclusion, the present study imposes a revision of our earlier interpretation of the adaptation observed in the ambiguous sound condition as reflecting exclusively visual recalibration. Exposure to a bimodal stimulus with an ambiguous auditory component would produce selective adaptation in parallel with recalibration. Due to the respective developmental courses of the two phenomena, recalibration would dominate the resulting AEs at early stages and be counterbalanced by selective adaptation later on.

Bertelson et al. (2003) noted that while both conflict-based recalibration and conflict-free sensory adaptation have been demonstrated in several other perceptual domains, their interaction had rarely been considered within the same experimental situation. A relevant case has been revealed by recent work concerning the influence of lexical context on auditory phoneme identification. In an important study, Samuel (2001) used the selective adaptation paradigm to demonstrate such lexical influences in the absence of contamination by post-perceptual adjust-

ments. His participants were exposed to repeated presentations of an ambiguous /s/–/ʃ/ sound in the context of either an /s/-final word (e.g. /bronchiti?/, from bronchitis), or an /ʃ/-final one (e.g. /demoli?/, from demolish). In post-tests involving identification of the ambiguous /s/–/ʃ/, fewer reports of a particular alternative were obtained after exposure to words favouring that alternative. Samuel concluded that the lexically induced phoneme had produced selective adaptation, in the same manner as an acoustically delivered one.

More recently, though, Norris, McQueen, and Cutler (2003) exposed listeners with similar materials, but instead of selective adaptation, they observed recalibration (or, in their terms, *perceptual learning*[1]). For instance, they replaced the final fricative (/f/ or /s/) from critical Dutch words by an ambiguous sound, intermediate between /f/ and /s/. Listeners heard this ambiguous utterance either in /f/-final words (e.g., /witlo?/, from witlof, chicory) or in /s/-final words (e.g., /naaldbo?/, from naaldbos, pine forest). Listeners who heard /?/ in /f/-final words were in subsequent testing more likely to report /f/, whereas those who heard /?/ in /s/-final words were more likely to report /s/. Thus, exposure to what seems to be very similar materials caused in one study (Samuel, 2001) selective adaptation, and the other (Norris et al., 2003) recalibration.

There are several differences between the two experiments that may explain the contradiction. For instance, Samuel used a straightforward selective adaptation method, the one used by Norris et al. involved less habitual procedures, like embedding the adapters in a larger number of neutral fillers. Our results however suggest that the critical factor may be the amount of exposure received by the participants. Norris et al. (2003) exposed their participants to just 20 inducing utterances, embedded in a single block among 180 fillers, while Samuel (2001) administered each utterance 768 times (24 blocks of 32 inducers, each followed by 8 post-tests, and no fillers). If the lexical effects taking place in these experiments involve, like the crossmodal effects studied here, an early phase dominated by recalibration (or perceptual learning) and a later phase dominated by selective adaptation, then a short adaptation phase (as in Norris et al., 2003) may reveal mainly recalibration, which, with the kind of massive exposure carried out by Samuel, would be overtaken by selective adaptation.

In his paper, Samuel (2001) reported only the mean adaptation effects over the whole experimental session. However, since series of post-tests were carried out after each of the 24 adaptation blocks, the data contained all the necessary information concerning the build-up course. Samuel has kindly made these data available to us. Fig. 3 shows AEs for successive adaptation blocks. Negative differences are observed for the clear majority of blocks posterior to block 3, showing the expected dominant role of selective adaptation. But a positive difference, possi-
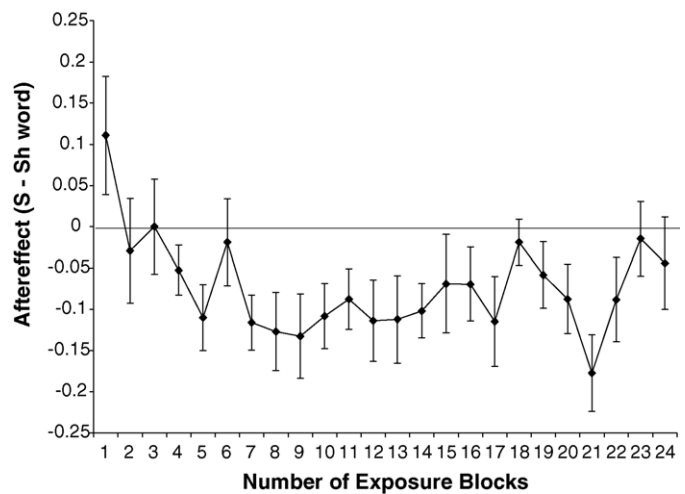

Fig. 3. Mean aftereffects, averaged across lexical contexts, as functions of exposure blocks, in the experiment of Samuel (2001). Exposure stimuli were words with final /s/ or /ʃ/, in which the final fricative had been replaced by an ambiguous intermediate sound (e.g. /bronchiti?/, from bronchitis), or /demoli?/, from demolish). Tests in which items from an 8-step /is/–/iʃ/continuum were categorized as /is/ or /iʃ/ were run after each block of 32 exposures. Aftereffects are measured by the proportion identifications consistent with the lexical inducers. (Data courtesy of Arthur Samuel.)

bly indicative of recalibration, obtains on block 1 (i.e. after 32 adapter presentations), and progressively gives way to negative ones on following blocks. Thus, the succession observed in our ambiguous sound condition of a pattern dominated by recalibration and of one dominated by selective adaptation is present in Samuel's lexical recalibration situation as well. It might occur generally during prolonged exposure to various sorts of conflict situations.

Acknowledgements

References

Bertelson, P. (1999). Ventriloquism: a case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal event* (pp. 347–362). Amsterdam: Elsevier.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 151–177). Oxford: Oxford University Press.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science*, *14*, 592–597.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Science*, *7*, 460–467.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

---

[1] *Recalibration* is the term classically used in the literature on multisensory perception, and on perceptual adaptation more generally, to designate conflict-based modifications in input-to-percept correspondences. *Perceptual learning*, which is currently gaining acceptance for similar usages in speech studies, misses the distinction between a modification in existing translating rules and the acquisition of new rules.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *The Quarterly Journal of Experimental Psychology*, *26*, 63–71.

Recanzone, G. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, *95*, 869–875.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, *30*, 309–314.

Rosenblum, L. D. (1994). How special is audiovisual speech integration? (Commentary on Radeau). *Current Psychology of Cognition*, *13*, 110–116.

Saldaňa, A. G., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society America*, *95*, 3658–3661.

Samuel, A. G. (1986). Red herring detectors and speech perception: in defence of selective adaptation. *Cognitive Psychology*, *18*, 452–499.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *18*, 452–499.

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 Communication*, *44*, 55–61.