

Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory

C. Colin^{a,*}, M. Radeau^{a,b}, A. Soquet^c, D. Demolin^c, F. Colin^d, P. Deltenre^{a,e}

^aResearch Unit in Cognitive Neurosciences, Université Libre de Bruxelles, CP 191, Av. F.D. Roosevelt, 50, B-1050, Brussels, Belgium

^bFNRS, Egmont Street, 5, B-1000, Brussels, Belgium

^cPhonology Laboratory, Université Libre de Bruxelles, CP 175, Av. F.D. Roosevelt, 50, B-1050, Brussels, Belgium

^dUniversité Libre de Bruxelles, Av. F.D. Roosevelt, 50, B-1050, Brussels, Belgium

^eEvoked Potentials Laboratory, Unit 16, Brugmann Hospital, Van Gehuchten Place, 4, B-1020, Brussels, Belgium

Accepted 30 January 2002

Abstract

Objectives: The McGurk–MacDonald illusory percept is obtained by dubbing an incongruent articulatory movement on an auditory phoneme. This type of audiovisual speech perception contributes to the assessment of theories of speech perception. The mismatch negativity (MMN) reflects the detection of a deviant stimulus within the auditory short-term memory and besides an acoustic component, possesses, under certain conditions, a phonetic one. The present study assessed the existence of an MMN evoked by McGurk–MacDonald percepts elicited by audiovisual stimuli with constant auditory components.

Methods: Cortical evoked potentials were recorded using the oddball paradigm on 8 adults in 3 experimental conditions: auditory alone, visual alone and audiovisual stimulation. Obtaining illusory percepts was confirmed in an additional psychophysical condition.

Results: The auditory deviant syllables and the audiovisual incongruent syllables elicited a significant MMN at F_z . In the visual condition, no negativity was observed either at F_z , or at O_z .

Conclusions: An MMN can be evoked by visual articulatory deviants, provided they are presented in a suitable auditory context leading to a phonetically significant interaction. The recording of an MMN elicited by illusory McGurk percepts suggests that audiovisual integration mechanisms in speech take place rather early during the perceptual processes. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Mismatch negativity; McGurk–MacDonald effect; Speech perception; Short-term memory

1. Introduction

Theories of speech perception can be categorized under two main classes: auditory or conventional theories and motor theories (Liberman and Mattingly, 1985; Mattingly and Liberman, 1988; Remez et al., 1994; Liberman, 1998). Auditory theories assume that general auditory processes are sufficient to explain the perception of speech (Pisoni, 1973; Cole and Scott, 1974; Kuhl, 1981; Massaro, 1987; Stevens, 1989). Motor theories invoke specialized processes using decoding rules based on knowledge of articulatory mechanisms (Liberman et al., 1967; Fowler, 1986). The so-called revised motor theory (Liberman and Mattingly, 1985) appears as the most unconventional and radical of its kind by proposing that the motor code used to understand speech is processed by a specialized perceptual module obeying the rules defined by Fodor (1983). According to

the revised motor theory, the object of speech perception is not the acoustic signal but the intended phonetic gestures of the speaker, so that the perception of an utterance is achieved by identifying a specific pattern of intended articulatory gestures. Under this theory, a specialized perceptual Fodorian module performs the automatic conversion of received acoustic signals into intended articulatory gestures, which lead to perceived phonetic structures. The specialized module is viewed as an innate piece of neural architecture resulting from the species evolution and performing an automatic key function in speech perception without any cognitive load. The theory assumes that the module is engaged whenever an acoustic stimulus can be interpreted as the result of linguistically significant articulatory gestures, and that it removes the information it processes from the input of the general acoustic perception mechanisms (Liberman and Mattingly, 1985; Mattingly and Liberman, 1988). This implies a serial link between the specialized phonetic

* Corresponding author. Tel.: +32-2-6504549; fax: +32-2-6502209.
E-mail address: ccolin@ulb.ac.be (C. Colin).

module and the general auditory mechanisms (Mattingly and Liberman, 1988).

Proponents of the revised motor theory have invoked the McGurk–MacDonald effect (McGurk and MacDonald, 1976) in order to support their views that speech perception is special. The McGurk–MacDonald effect is obtained as follows: via an audiovisual system, a subject is presented with a speaker who articulates a syllable different from the one recorded on the audio track. For example, the audio signal presents /gi/ but the speaker silently articulates /bi/. The resulting percept is /bgi/, a combination of the incongruent auditory and visual information. Other incongruences, like a visual /gi/ dubbed on an auditory /bi/ lead to a fused percept /di/ (see Massaro, 1987, 1998; Hardison, 1996 for reviews on the McGurk–MacDonald effect). The McGurk–MacDonald effect demonstrates the convergence of acoustic and visual information on the phonetic perception. The effect is obligatory and irrepressible, even when the subject has full knowledge of the actual stimuli (Summerfield and McGrath, 1984; Rosenblum and Saldaña, 1996). For the proponents of the revised motor theory, the percept is unambiguously phonetic, not bimodal, and this brings support to the assumption that speech perception is neither auditory nor visual, but results from the early conversion of received visual and acoustic information into intended articulatory gestures. According to this view, speech perception is heteromorphic since the perceived objects possess dimensionalities radically different from those of the proximal stimulus (Mattingly and Liberman, 1988).

The fuzzy logic model of perception (FLMP) provides quite a different view about the McGurk–MacDonald effect (Massaro, 1987). For this model, which belongs to the auditory class of speech perception theories, the illusory phonetic percepts are not heteromorphic and result from a rather late integration process that follows independent, parallel evaluation of both input modalities (Massaro, 1987).

Since its discovery, the McGurk–MacDonald effect has been extensively applied to the study of audiovisual speech perception. The results from those studies have been used to assess and improve current theories of speech perception (see Green, 1998 for a review).

The mismatch negativity (MMN) is an event-related potential (ERP), which indexes the pre-attentive, automatic detection of deviance of a rarely occurring stimulus with respect to frequent, standard background stimuli (for reviews, see Näätänen, 1992; Ritter et al., 1995; Lang et al., 1995; Näätänen and Alho, 1997; Schröger, 1997). There are a lot of studies leading to believe that the MMN is the outcome of an automatic comparison process between a new, deviant stimulus and the memory trace formed by the sensory representation of the standard stimuli within the short-term memory (STM) (Novak et al., 1990; Näätänen, 1990, 1992; Cowan et al., 1993; Cowan, 1995; Näätänen and Alho, 1995). It is not completely certain whether the MMN is specific or not to the auditory modality. Initial experiments

seeking it in the visual modality concluded to its absence (Nyman et al., 1990; Czigler and Csibra, 1990) but later research suggested that an MMN-like phenomenon might be present in the somatosensory modality (Kekoni et al., 1996), in the chemosensory modality (Pause and Krauel, 2000) and in the visual modality (Cammann, 1990; Woods et al., 1992; Alho et al., 1992; Tales et al., 1999). If a visual MMN exists, current data suggest that it exhibits a restricted occipito-temporal scalp distribution consistent with an origin in modality specific sensory cortex (Woods et al., 1992; Alho et al., 1992), whereas the auditory MMN has a fronto-central distribution and is known to have a major source in the auditory cortex, with a secondary contribution from the frontal lobe (Giard et al., 1990; Alho, 1995). Another significant difference with the auditory MMN could be a higher threshold, the visual MMN-like response requiring larger contrasts above the discriminative threshold to be elicited (Alho et al., 1992). The auditory MMN often inverts in polarity at scalp locations below the sylvian fissure, typically between the F_z and M_1 or M_2 electrodes (Alho, 1995; Novak et al., 1990). Some types of auditory contrast do, however, evoke an MMN that does not invert its polarity across the sylvian fissure (Sussman et al., 1998a,b), thus suggesting different cerebral generators. The MMN is best isolated in passive conditions in which the subject does not pay attention to the stimuli, otherwise one is likely to record a mixed waveform made up of the pure MMN mixed with the N2–P3 complex typical of active conditions (Lang et al., 1995; Näätänen, 1995).

The MMN has already proved to be a very interesting tool to investigate different aspects of speech perception in normal as well as pathological populations (Aaltonen et al., 1987; Sams et al., 1990; Kraus et al., 1992, 1995, 1996; Ponton and Don, 1995).

Several studies have used the phonetic category boundary effect in order to record an MMN that would index phonetic rather than acoustic deviance (Aaltonen et al., 1992; Maiste et al., 1995; Sams et al., 1990; Sharma et al., 1993; Dehaene-Lambertz, 1997). Phonetic categories are a consequence of categorical perception, which is a major, although not specific feature of speech perception (Liberman et al., 1957; Cutting and Rosner, 1974; Miller et al., 1976). Unfortunately, these studies either did not show a phonetic effect or were unable to unambiguously separate the effects of acoustic from those of phonetic deviance (see Winkler et al., 1999 for a discussion of these studies). Recently, however, investigations capitalizing on the difference between phonetic category borders in Finnish and Estonian and using vowels that are acoustically simpler, rather than the traditional consonant-vowel syllables, showed that a language-specific phonetic representation of deviant vowels also activates the MMN discriminative process (Näätänen et al., 1997; Cheour et al., 1998). A further elaboration of this type of study used Hungarian and Finnish languages. This allowed the use of two pairs of synthesized isolated vowels, the former representing an across-vowel category in Hungarian and a within-category in Finnish whereas the latter had the reverse role in

both languages. The MMN amplitude was larger in across-than within-category contrasts; in other words, speakers of each language had a larger MMN to the contrast that crossed a vowel boundary in their native language than to the contrast between two identically categorized vowels (Winkler et al., 1999). The conclusion of this latter study was that the MMNs elicited by auditory (sensory) and phonetic (categorical) deviations can proceed in parallel suggesting that these two forms of stimulus representations coexist in the brain. Moreover, in the Finnish-speaking group, the Finnish contrast elicited an MMN with two successive peaks, whereas in the Hungarian group, the same Finnish contrast elicited a simple peak MMN at the latency of the second peak of the Finnish subjects. This suggests that the earlier peak that occurred in the Finnish group was phonetically evoked, a finding compatible with several postulates of the revised motor theory.

The present study was designed as an exploratory investigation in order to test the hypothesis that rare incongruent visual stimuli dubbed onto constant auditory syllables would evoke an MMN by creating a deviant phonetic percept through the McGurk–MacDonald effect. If it succeeded, such a paradigm should be able to isolate the sought-for pure phonetic MMN, since the auditory stimuli are invariant and the deviant-incongruent vs. standard-congruent visual stimuli should evoke either no MMN or a sensory-specific one, easily distinguishable by its posterior localization. The presence of an MMN evoked by the McGurk–MacDonald effect was deemed fairly probable since the same stimulation paradigm has already been shown to evoke differential magnetic field responses considered to be a probable equivalent to the electrical MMN (Sams et al., 1991). Moreover, there exist two preliminary reports that claim the obtention of an MMN evoked by the McGurk–MacDonald effect (McPherson and Andrews, 1997; Bernstein et al., 2001).

2. Methods

2.1. Subjects

2.1.1. Pre-selection from the original pool

Since our resources in terms of subject number and retrievable recording sessions were limited, pilot recordings of the MMN evoked by pure tones were performed in order to limit the extensive complete recording sessions to subjects who proved to yield robust MMNs. A pure tone frequency contrast was investigated. Frequent (5700 presentations) and rare (600 presentations) stimuli were contrasted by a comfortable frequency separation (1000 and 1250 Hz). The pure tone stimuli lasted 400 ms (20 onset–offset ramps). The selection was based on a visual judgment of the final waveform quality and was performed within an initial pool of 21 subjects. Since the MMN is known to suffer from a high interindividual variability (Lang et al., 1995; Uwer and von

Suchodoletz, 2000) and the basic question investigated concerned the existence or not of an MMN evoked through the McGurk–MacDonald effect, the pre-selection procedure aimed at avoiding contamination of the results by the random inclusion of a high proportion of individual subjects with intrinsically poor MMN from a relatively small group.

2.1.2. Selected subjects

Three subjects did not yield any recognizable MMN, 10 others showed either doubtful ($<1 \mu\text{V}$) MMNs or poor recordings due to movement artefacts.

Eight right-handed normal subjects (4 males, 4 females) aged 17–62 years were therefore selected to participate in the complete experiment as paid volunteers. They were in good health, had normal auditory function (pure-tone air-conducted hearing thresholds at 20 dB or better for every octave frequency between 125 and 4000 Hz) and normal or corrected-to-normal vision. All were native speakers of French. They gave their informed consent to participate after the details of the procedure had been explained to them. The experimental protocol has been approved by the ethical committees of Brussels University and of Brugmann Hospital where the neurophysiological recordings took place.

2.2. Experiments

After the initial pre-selection procedure based on individual MMN quality, five experiments were conducted in the present investigation.

2.2.1. Experiments 1–3

A high-speed video camera was configured to capture gray scale images (240×192 pixels) at a rate of 125 images per second. Audio signal and trigger pulses corresponding to image acquisition were recorded simultaneously on a digital audio tape. Pulses were used to post-synchronize the audio signal with the video data. Each stimulus consisted of one acoustic and one video track. Both tracks were synchronized on the burst of the stop consonant measured on the acoustic signal. The acoustic track was down-sampled to 22.05 kHz and the video track to 25 images per second. The stimuli were played back by means of the iShell software version 1.2 (<http://www.tribeworks.com>)

The video screen (standard 17 inches computer color video monitor) was placed 100 cm in front of the subjects who sat comfortably in an armchair fitted with a back headrest. The acoustic part of the stimuli was delivered at an overall intensity of 60 dB sound pressure level through a loudspeaker positioned immediately below the video screen.

Experiment 1 investigated the MMNs evoked by purely auditory stimuli which were the syllables recorded on the audio track of the audiovisual stimulation sequence with the video screen switched off.

Experiment 2 investigated the MMNs evoked by purely

Table 1
Stimuli pairs

Occurrence	Auditory		Visual		Audiovisual	
	/bi/	/gi/	/bi/	/gi/	A/bi/V/bi/	A/gi/V/gi/
Frequent	/bi/	/gi/	/bi/	/gi/	A/bi/V/bi/	A/gi/V/gi/
Rare (10.5%)	/gi/	/bi/	/gi/	/bi/	A/bi/V/gi/	A/gi/V/bi/

visual stimuli that consisted of the audiovisual sequences played with the sound channel turned off.

Experiment 3 investigated the MMNs evoked by the combined audiovisual stimuli with congruent stimuli as standards and incongruent as deviants.

For each of these 3 experiments (auditory alone, visual alone and audiovisual), two symmetrical contrasts were built. Table 1 presents the stimuli pairs used in each condition.

In order to minimize the risk of MMN habituation reported to occur when recordings exceed 10 min (Cheour et al., 2000), stimuli were presented in short sessions containing 570 standard and 60 deviant stimuli (deviant probability of occurrence: 10.5%) delivered at a rate of 0.909 Hz. The stimulus onset asynchrony measured from the start of the earliest of the stimuli component was 1100 ms. Stimulus duration was variable according to the components making up the whole stimulus. The acoustic component always lasted 341 ms, the visual components lasted 582 ms for /gi/ and 372 ms for /bi/. The number of standard stimuli separating two deviant ones was randomly varied between 6 and 11. Subjects were free to take any type of break they wished between short sessions. A complete MMN data bank for a given pair of standard-deviant stimuli was made up of 5 short sessions, in order to include 300 deviant stimuli before data analysis. A maximum of 10 short sessions were recorded in the same day.

The presentation order of stimuli pairs was randomized between short sessions with the additional constraint that the same contrast was never presented more than once in immediate succession, and that the same modality was never presented more than twice in immediate succession.

2.2.2. Experiment 4

Experiment 4 explored the polarity reversal behavior of the audiovisual MMN. It was performed on one single subject who agreed to participate in extensive recordings in order to achieve a final signal-to-noise ratio (SNR) sufficiently high to warrant demonstration of a reversal should one be present.

Four types of contrasts were investigated.

A pure tone frequency contrast: 1000 Hz (standard) vs. 1250 Hz (deviant) stimuli that were identical to those used in the pre-selection procedure.

An auditory spatial localization contrast: the auditory syllable /bi/ delivered by the loudspeaker positioned immediately below the video screen (standard) vs. delivered by a loudspeaker located at 60° to the right of the

subject (deviant). The /bi/ syllable was that used in experiment 1.

An auditory syllable contrast: /gi/ standard vs. /bi/ deviant, using the same syllables as in experiment 1.

An audiovisual syllable contrast evoking the McGurk–MacDonald illusion : A/gi/ V/gi/ vs. A/gi/ vs. V /bi/, as in experiment 3.

As in experiments 1–3, stimuli were presented in short sessions containing 570 standard and 60 deviant stimuli delivered at a rate of 0.909 Hz. The complete MMN data bank for a given pair of standard–deviant stimuli was made up of 21 short sessions for the first 3 contrasts and of 28 sessions for the audiovisual contrast. The same contrast was never presented more than twice in immediate succession. This experiment first sought to define the minimal SNR that MMNs known to invert their polarity at mastoid electrodes should reach in order to be able to objectively demonstrate such an inversion on the basis of a statistical method applied to mastoid recordings. For each of the first 3 contrasts used (pure tones, auditory alone syllables, localization), grand averages were progressively built-up by successive addition of the subaverages collected within each short session. For each subaverage increment, the grand average SNR was computed for the MMN waveform and Guthrie's test (Guthrie and Buchwald, 1991) was applied to the mastoid recordings to test for a significant positive deflection. Once the minimal SNR established, the MMN evoked by the McGurk–MacDonald illusion was averaged up to a SNR well above this value in order to be confident that an absence of polarity reversal could not be due to poor SNR.

2.2.3. Experiment 5

Experiment 5 was purely psychophysical and measured the percentage of McGurk–MacDonald illusions provided by the incongruent audiovisual stimulation pairs. For each contrast pair, the percentage of illusions was computed as the ratio between the number of reported illusory (fusion or combination) percepts and the total number of trials (12 per contrast pair). Subjects indicated what they had perceived by selecting one of 4 possible answers (/bi/, /gi/, /di/ or /bgi/) on a multiple choice sheet.

2.2.4. Control of the attentional state of the subjects

In order to ascertain the inattentive condition requested for uncontaminated MMN recording, subjects were requested to perform a tactile discrimination task in experiments 1–4. This allowed for the necessary visual fixation of the video screen. Vibratory stimuli were applied to the subjects' right wrist by acoustically coupling the output of an auditory stimulator with an acoustico-mechanical transducer (mini-fonator, Siemens) normally used for vibrotactile stimulation of profoundly deaf subjects. Four types of tasks were constructed: 3 involved discrimination of a single parameter value (amplitude, frequency or duration) between the middle stimulus and the two flanking ones.

These 3 stimuli made up a presentation interval. The fourth task involved the detection of a gap in a single stimulus filling up each presentation interval. A trial consisted in gradually varying the target parameter from one presentation interval to the next one, until the subject signaled that he had perceived a difference by hitting a key, whereafter a new trial was immediately started. From trial to trial, the target parameters were randomly modified in type (amplitude, frequency, duration or gap), as well as in ascending or descending order and in step size. The tasks and the low-frequency acoustic stimuli feeding the vibrating transducer were programmed using Tucker–Davis Technologies hardware (System II) and software (Siggen and Sigplay).

2.3. Electrical activity recording (experiments 1–4)

Brain electrical activity was recorded with Ag–AgCl electrodes from F_z , O_z , C_3 , C_4 , (10–20 system) as well as from left and right mastoids (M_1 , M_2), all referred to the tip of the nose. Horizontal and vertical eye movements were monitored using two bipolar recordings: one between each outer eye canthus and one between a supraorbital electrode and an electrode just below the lower lid on the right side. After amplification (10,000 times for brain activity channels and 5000 times for eye movement channels) and filtering (0.1–70 Hz), the input signals were digitized with a sampling rate of 1138 Hz and stored on the computer disk for off-line averaging.

2.4. MMN computation and measurement

Data averaging was performed with an InstEP system. Recording epochs with a total duration of 900 ms were averaged separately for the standard and deviant stimuli. A pre-stimulus baseline of variable duration made up the early part of the epochs: when a visual component was present, the baseline duration (prior to the onset of articulatory movement) was, respectively, 10 for /gi/ and 134 ms for /bi/; it was 165 ms for auditory alone stimuli. Rejection of artefacted sweeps from the final averages was performed on the basis of a $\pm 100 \mu\text{V}$ criterion for all channels. Averaged waveforms were converted into ASCII format and exported to a spreadsheet program for further analysis and plotting. The MMN was computed as the differential waveform obtained by subtracting the potential evoked by frequent stimuli from the one evoked by deviant stimuli.

For each of the contrasts listed in Table 1, the presence of an MMN at F_z and O_z locations was objectively ascertained from the grand average computed across the 8 subjects according to the method described by Guthrie and Buchwald (1991). This method is based on a succession of paired one-tailed t tests that compare the point-to-point amplitudes of the two across-subjects grand average waveforms (the one evoked by frequent stimuli and the other by rare ones), using individual waveforms ($N = 8$) as a data set. The heart of the method relies on the computation of the minimum number of consecutive, adjacent individual differences

that should be found significant at the 5% level, in order to be above the incidence of randomly occurring consecutive significant results due to multiple t tests in the absence of any genuine difference between the two sets. The statistical computations including the autocorrelation estimates needed to use Guthrie and Buchwald's method were computed using Pascal routines that are widely available (Press et al., 1990). The original data points were first digitally low-pass filtered (0.1–30 Hz) then decimated in order to achieve a final digital temporal resolution of 6.16 ms compatible with Guthrie and Buchwald's table. The autocorrelation values attained 0.9 for all contrasts at both locations. According to Guthrie and Buchwald's table, this indicates that 12 successively significant point-to-point t tests, spanning a 67 ms temporal threshold are required in order to define a statistically significant MMN at the 5% level (Guthrie and Buchwald, 1991).

MMN latency was defined as that of the most negative point within the period of consecutive significant t tests above Guthrie and Buchwald's temporal threshold. MMN amplitude was defined as the absolute amplitude of the most negative point within the period of consecutive significant t tests above Guthrie and Buchwald's temporal threshold.

MMN duration was defined as the fraction of the recording epoch during which significant consecutive t tests were obtained above Guthrie and Buchwald's temporal threshold. When more than one period of significant t tests was obtained, the MMN was considered to be made up of several components.

2.5. SNR measurement (experiment 4)

The SNR of the 4 different MMNs was computed by dividing the root-mean-square amplitude of the post-baseline data points by that of the baseline data points (Arnold, 1985; Boston, 1989; Deltenre and Mansbach, 1993).

3. Results

3.1. Experiment 1 (auditory presentation alone)

Fig. 1 illustrates the waveforms evoked by the two contrasts delivered in the auditory alone modality.

The standard and deviant stimuli elicited obvious F_z components, morphologically similar to the classical exogenous P_1 – N_1 – P_2 components of the long-latency "Vertex" auditory potentials (Davis, 1976). When measured from the consonant burst onset on the waveform evoked by the standard /bi/ stimulus, the respective latencies of these 3 components were 44, 88 and 172 ms, suggesting that they were indeed synchronized on the consonant burst.

In order to provide the reader with a detailed view of the temporal relations between the acoustic–phonetic stimuli and the ERP they evoke, Fig. 2 shows, on the same time scale, the temporal pressure wave of the syllable /bi/ with its spectrographic representation and the associated exogenous evoked potential. The spectrogram plots the variations of

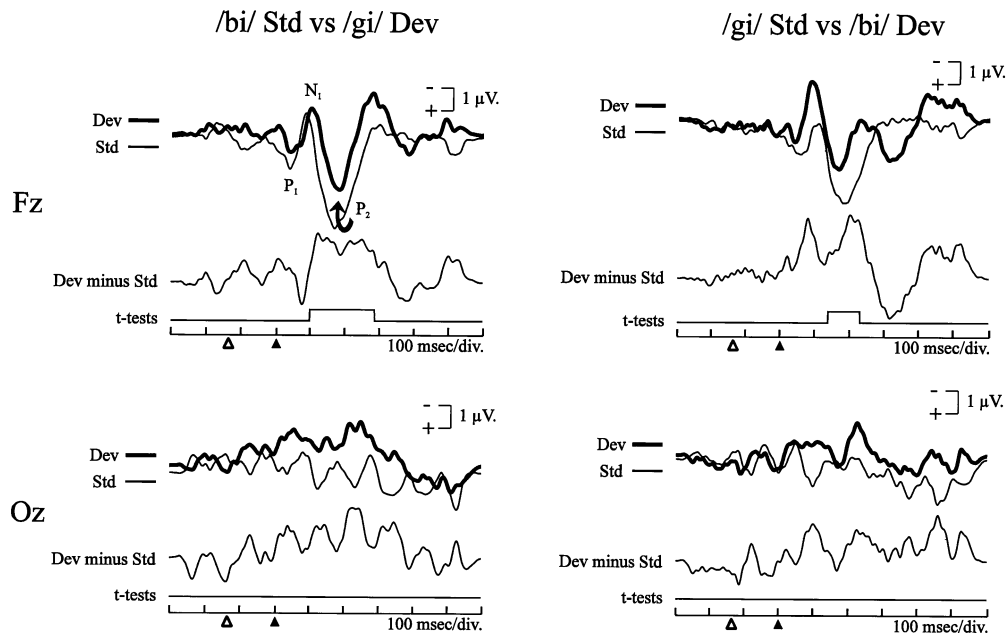


Fig. 1. Auditory alone presentation. In this and subsequent figures, the potentials evoked by the standard (thin line) and deviant (thick line) stimuli are superimposed and represent grand averages across all subjects. The classical exogenous P_1 , N_1 and P_2 components are readily identified. The derived waveform obtained by subtracting the standard response from the deviant one is plotted below the exogenous waveforms. The lowest trace illustrates the result of the statistical testing on a binary mode: the level is raised during each period of consecutive significant t tests exceeding Guthrie's temporal threshold. The black triangles indicate the temporal position of the deviant consonant burst, the open triangle indicates deviant voicing onset. Both contrasts evoke an MMN at F_z location, whereas none is found at O_z .

the acoustic signal frequency content across time. This helps to confirm that the obvious N_1 – P_2 component that dominates the ERP waveform is synchronized on the consonant burst explosion.

The two acoustic-phonetic contrasts (/bi/ standard vs. /gi/ deviant and /gi/ standard vs. /bi/ deviant: Fig. 1) evoked clear-cut MMNs at F_z but not at O_z . When /gi/ was the deviant stimulus, the MMN had a rather long-duration of 185 ms which covered the time slot extending between the exogenous N_1 peak up to the end of the P_2 peak. This fairly long MMN contained two negative peaks with respective latencies of 125 and 251 ms. When /bi/ was deviant, the MMN was shorter (90 ms), peaked at 200 ms and covered most of the exogenous P_2 . A first negative peak with a short latency (78 ms), visually appealing as a candidate for a very early MMN failed to fulfill the statistical criterion requested to be identified as such. Both auditory contrasts evoked an MMN that inverted its polarity between the F_z and M_1 as well as M_2 electrodes (see experiment 4 and Fig. 5 for further details).

3.2. Experiment 2 (visual presentation alone)

As shown in Fig. 3, no MMN could be detected either at F_z or at O_z .

3.3. Experiment 3 (audiovisual presentation leading to the McGurk–MacDonald effect)

The waveforms evoked at F_z by the standard and deviant stimuli exhibited components, which were morphologically

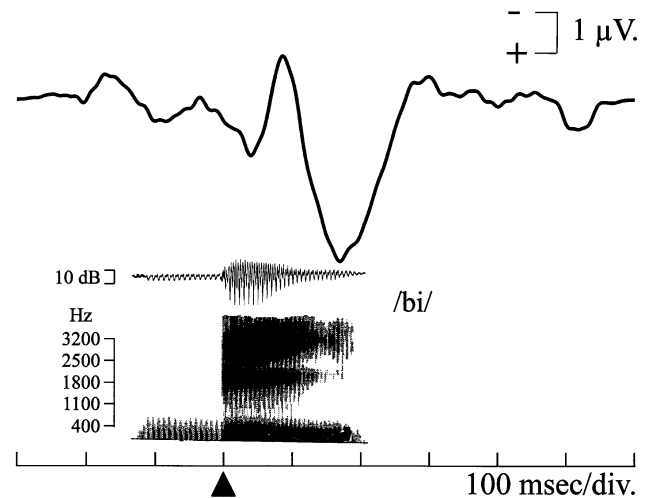


Fig. 2. Temporal relationships between the features of an acoustic syllabic stimulus (/bi/) presented alone and its ERP. Upper trace: ERP recorded from F_z . Middle trace: acoustic pressure waveform as measured by the microphone used to record the utterance before digitizing. Lower trace: spectrogram of the syllable. The spectrogram displays frequencies (ordinate) over time (abscissa). At each point in time, the intensity of a given frequency component is coded by the relative darkness of the tracing. The first acoustic event is the low-frequency voicing bar, corresponding to the onset of laryngeal vibrations, and starts 165 ms after sweep onset. This event is followed 135 ms later (at 300 ms; black triangle) by a burst of acoustic energy at higher frequencies, which is due to the release of the lip closure. The burst is, in turn, followed by rapid frequency transitions towards the long duration stable frequency components (formants) that characterize the vowel.

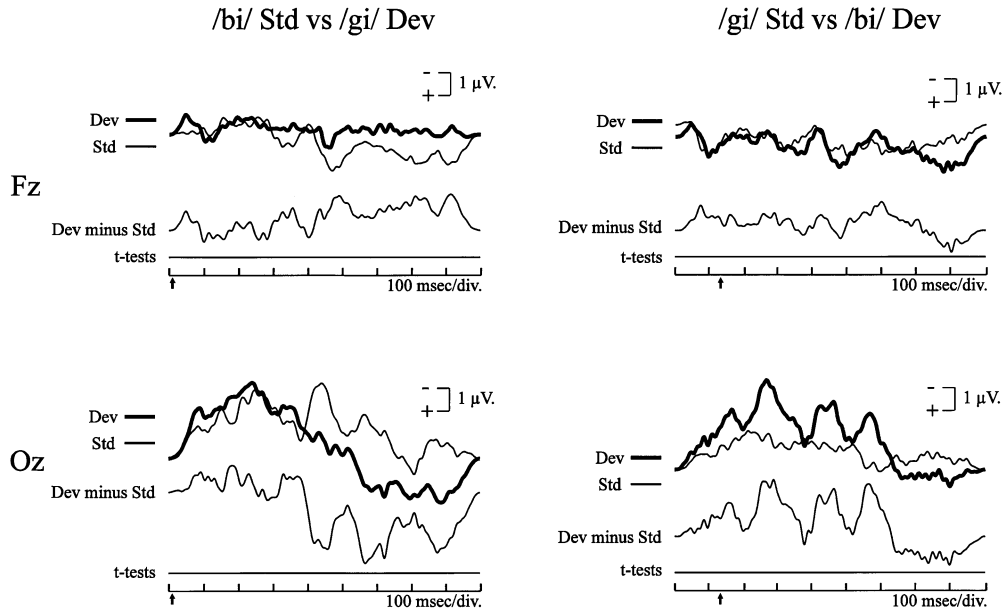


Fig. 3. Visual alone presentation. The arrows indicate the onset of articulatory movement of the deviant stimuli. No MMN can be detected whatever the contrast and the scalp location.

very similar to the exogenous P₁, N₁ and P₂ waves obtained in the auditory alone condition. The components evoked by the standard /bi/ stimulus had latencies of 45, 94 and 167 ms with respect to consonant burst.

Both phonetic contrasts created by the McGurk–MacDonald effect evoked an MMN at F_z, and not at O_z as illustrated in Fig. 4.

When /gi/ was deviant, the MMN appeared to contain 3 components:

An early component with a duration of 119 ms and covering most of the P₁ as well as the ascending limb and peak of the N₁ components evoked by the deviant and standard stimuli.

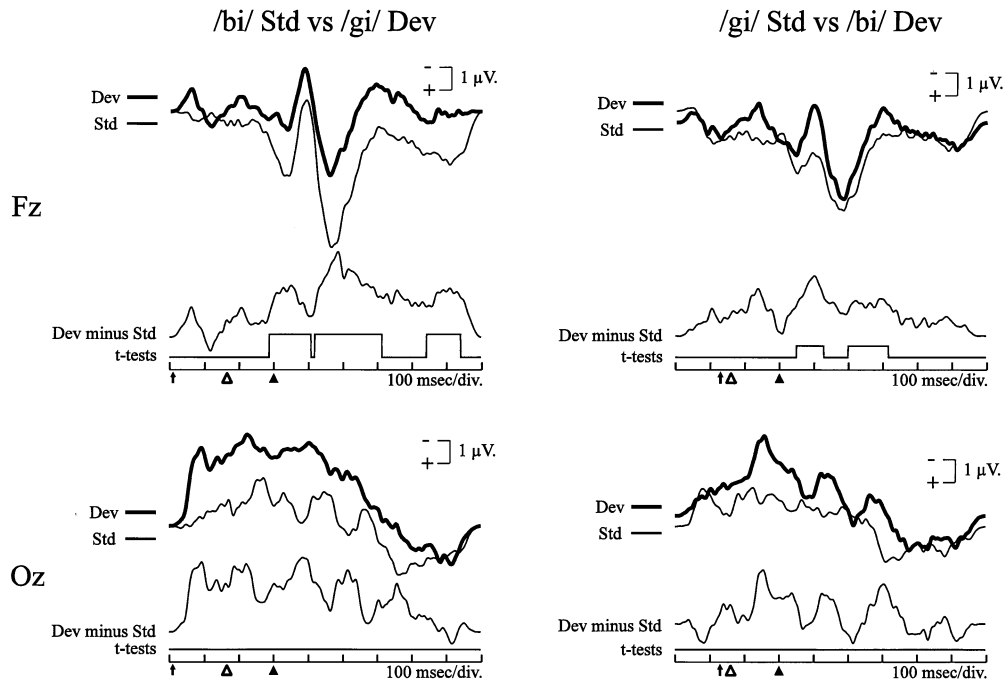


Fig. 4. Audiovisual presentation giving rise to the McGurk illusion. Arrows and triangles indicate the same stimulus temporal features as in previous figures. Both contrasts elicit MMNs at F_z only. When /gi/ is deviant, 3 components are identified. The first two cover most of the exogenous waves, the third one is very late. When /bi/ is deviant, two components are identified: one over the N₁ time slot and the other over the second half of P₂.

An intermediate component with a duration of 193 ms which covered the P₂ time slot of the deviant and standard waveforms.

A very late component with a duration of 98 ms.

When /bi/ was deviant, an MMN made up of two components was obtained:

An early component with a duration of 78 ms, which covered the ascending limb and peak time slot of the N₁ exogenous components.

A later component starting just after the peak of the exogenous P₂ and lasting for 116 ms.

None of the two audiovisual contrasts did evoke a significant MMN that inverted its polarity between the F_z and M₁ or M₂ electrodes. This peculiar point was further studied in experiment 4.

3.4. Statistical comparison between audiovisual and visual alone conditions

The difference between the audiovisual and visual alone conditions was further tested by performing a statistical comparison of both the peak and mean amplitudes measured during the time periods for which Guthrie's test indicated the presence of a significant MMN in the audiovisual condition. Since these time periods were different between the two phonemic contrasts, two-by-two *t* tests were preferred to an analysis of variance (ANOVA) that would have included non-significant portions of the waveforms. Paired *t* test between the audiovisual and visual conditions were thus computed for each MMN time period of both contrasts.

T tests performed on mean amplitudes supported the presence of an audiovisual double-peaked MMN for the deviant /bi/ (earliest MMN period: $t(7) = 3.02$, $P < 0.01$; latest MMN period: $t(7) = 1.86$, $P < 0.05$), whereas those performed on peak amplitudes only confirmed the presence of the earliest MMN period (earliest MMN period: $t(7) = 2.79$, $P < 0.01$; latest MMN period: $t(7) = 0.73$, $P < 0.50$). For the deviant /gi/, *t* tests performed on mean amplitudes only supported the significance of the earliest MMN period ($t(7) = 1.85$, $P < 0.05$; middle latency MMN period: $t(7) = 1.59$, $P < 0.10$; latest MMN period: $t(7) = 0.56$, $P < 0.30$), whereas those run on peak amplitudes confirmed the presence of the 3 peaks (earliest MMN period: $t(7) = 3.15$, $P < 0.01$; middle latency MMN period: $t(7) = 3.56$, $P < 0.01$; latest MMN period: $t(7) = 2.62$, $P < 0.05$).

3.5. Parametric overview of experiments 1–3

Table 2 summarizes and compares the parameters resulting from the objective detection of the MMNs evoked in the two conditions for which an MMN was detected at F_z.

3.6. Experiment 4 (polarity reversal of the McGurk–MacDonald MMN)

Fig. 5 illustrates the results of experiment 4 and shows that a clear-cut phase reversal, mirroring the F_z MMN, was obtained for the pure tones frequency, auditory phonemic and spatial localization contrasts. The minimal SNR values (computed on the F_z differential waveform) at which Guthrie's test indicated a significant positivity at both right and left mastoid sites were, respectively, 1.62, 2.35 and 1.87. The MMN evoked by the McGurk–MacDonald illusion was averaged until its SNR reached 2.80, thus well above the minimal SNR that allowed statistical detection of polarity reversal for the three other contrasts (Fig. 5). The individual subject used in experiment 4 (Fig. 5) exhibited an MMN that was significant between 395 and 462 ms at the frontal site. Although visual inspection of the tracings reveals that the first half of the MMN wave is mirrored by a positive deflection at both mastoids, Guthrie's test failed to reveal a significant positivity at either mastoid electrode despite the enhanced SNR incorporating 1680 deviant stimuli. Further improvement of the mastoid waveform obtained by averaging the right and left tracings still did not yield a statistically significant positivity. It must therefore be concluded that a polarity reversal of the McGurk MMN is either absent or much less expressed than what is seen with auditory alone contrasts.

3.7. Experiment 5 (psychophysical McGurk–MacDonald effect)

Audiovisual incongruent stimuli gave rise to combination-type illusions in 74% of the trials and to fusion-type illusions in 66% of the cases.

4. Discussion

The main result of this study consists in the demonstration of an MMN evoked by illusory McGurk–MacDonald percepts. Although neither experiments 3 and 4 were able to demonstrate a significant polarity reversal between F_z and mastoid sites, there is no reason to refute the identification of the significant negative waveforms obtained in the audiovisual experiment as MMNs. Although it is true that most types of auditory contrasts evoke MMNs that invert their polarity, some clearly do not or do so only marginally (Sussman et al., 1998a,b). It follows that the definition of the MMN does not imply polarity reversal across the sylvian fissure, as further evidenced in the literature about the suspected visual, somatosensory and chemosensory MMNs (Cammann, 1990; Alho et al., 1992; Woods et al., 1992; Kekoni et al., 1996; Tales et al., 1999; Pause and Krauel, 2000). It has furthermore been shown that different types of auditory contrasts evoke MMNs with different topographical distributions (Giard et al., 1995; Paavilainen et al., 1991; Dehaene-Lambertz, 2000). The present results

Table 2
F_z recordings^a

MMN at F _z electrode																
	Onset latencies (ms)						Duration (ms)		Peak latencies (ms)						Peak amplitude (μV)	
	Articulatory gesture		Voicing		Consonant burst				Articulatory gesture		Voicing		Consonant burst			
	/gi/ dev	/bi/ dev	/gi/ dev	/bi/ dev	/gi/ dev	/bi/ dev			/gi/ dev	/bi/ dev	/gi/ dev	/bi/ dev	/gi/ dev	/bi/ dev		
A	NA	NA	276.9	312.9	102.8	138.9	185.4	89.6	NA	NA	298.8 (424.5)	373.5	124.8 (250.5)	199.5	-2.3 (2.0)	-2.7
AV 1	290.0	248.7	165.2	226.8	-8.8	52.7	118.7	78.2	351.6	299.7	226.8	277.7	52.7	103,711	-2.3	-2.1
AV 2	420.1	398.1	295.3	376.2	121.3	202.1	193.4	116.0	486.0	490.4	361.2	468.5	187.2	294.4	-3.7	-1.1
AV 3	743.6	-	618.8	-	444.7	-	97.6	-	809.5	-	684.7	-	510.6	-	-2.1	-

^a MMN parameters (onset latency, duration, peak latency and peak amplitude) for both auditory (A) contrasts (/gi/ deviant and /bi/ deviant) and for each significant MMN component (AV 1, AV 2, AV 3) of both audiovisual contrasts (visual /gi/ deviant and visual /bi/ deviant). The values of the MMN parameters have been computed as a function of several time points: the onset of the articulatory gesture (articulatory gesture), the onset of voicing and the onset of the consonant burst. The letters NA (non-applicable) mean that the values could not have been computed (i.e.: no articulatory gesture for the auditory condition). For the /gi/deviant of the auditory condition, two peaks of about the same amplitude were observed. The peak latency and the amplitude of the latest peak are indicated between brackets.

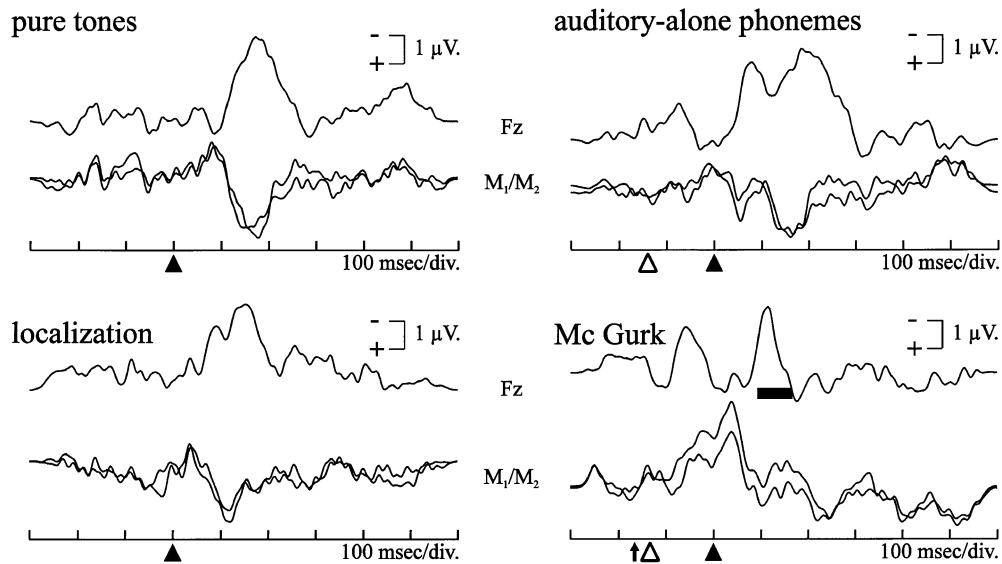


Fig. 5. Comparison of the polarity reversal behavior between F_z and mastoid sites for 3 types of auditory alone contrasts and for the McGurk contrast.

thus suggest that the MMN evoked by the McGurk–MacDonald illusory percepts originates from other generators than those involved with primary acoustic dimensions.

The existence of an MMN specific to the audiovisual modality is demonstrated by the results of the statistical comparison between the audiovisual and visual alone conditions. For both parameters that were taken into account (mean amplitude and peak amplitude), the early MMN component of the two audiovisual contrasts reached statistical significance. The middle latency and late MMN components were only significant on the basis of one of the two parameters analysis.

Since the audiovisual stimuli were devoid of any physical auditory contrasts and since the efficient visual contrasts did not evoke any MMN when presented alone, it appears that, within the conditions of the present experiments, the visual stimuli need a suitable auditory context in order to trigger an MMN associated with the presentation of an illusory deviant percept.

Since the MMN evoked by the McGurk–MacDonald effect is neither visual nor auditory and reflects the detection of a deviant phonetic percept, it is justified to propose that it results from an automatic, pre-cognitive comparison between STM phonetic traces. There exists no current agreement about the processing level (i.e. early or late with respect to mapping to phonetic prototypes) at which the audiovisual integration leading to McGurk–MacDonald illusory percepts occur (Green, 1998). The present data do not prove either that the STM trace has already been phonetically categorized at the point in time when it triggers the MMN, they at least indicate that the involved trace possesses an intrinsic phonetic value whatever its current processing stage.

Previous studies dealing with phonetic MMNs reported rather short latencies compatible with the hypothesis of

early processing of the phonetic dimension within the STM (Winkler et al., 1999). The onset latencies observed in the present study were also rather short, at least shorter than those of the MMNs evoked by similar pure auditory contrasts in the auditory alone conditions (Experiment 1). One must remain cautious when dealing with MMN latencies because the latency-reference point in time is defined by the moment at which the standard and deviant stimuli begin to diverge and this can occur much later than at deviant stimulus onset. Although visual deviance of articulatory movements is less straightforward to define than for auditory signals, it is probable that the visual deviant stimuli used in the present study diverged early, thus possibly contributing to the rather short MMN latencies observed. Although the very existence of an MMN evoked by a McGurk–MacDonald percept must be considered as favoring an early pre-conscious processing mechanism leading to the illusion, we need further experiments with parametric control of the timing of the visual and auditory stimuli to better define the moment at which the illusory deviance is detected.

The demonstration of the existence of an electrical MMN evoked by McGurk–MacDonald percepts is particularly welcome since it provides a relatively easy and hopefully universal way to probe STM function with pure phonetic traces. Although it is not language-independent, the McGurk–MacDonald effect has been demonstrated in many languages (Burnham, 1998) thus providing much more widespread possibilities to assess isolated phonetic percepts than by resorting to carefully selected auditory and phonetic contrasts from peculiar languages like Hungarian and Finnish (Winkler et al., 1999). Since both the McGurk–MacDonald effect and the MMN are elicitable across a wide range of ages including very young babies (Rosenblum et al., 1997; Burnham and Dodd, 1996; Cheour-

Luhtanen et al., 1995), their successful combination applied to the assessment of phonetic perception appears very promising regarding developmental issues in general, as well as in the study of various language acquisition problems which typically occur in pediatric age groups.

Acknowledgements

This work was financially supported by grants EC9011CL0780 and EC9011000007 from Université Libre de Bruxelles and Fonds National de la Recherche Scientifique to M. Radeau, the 'Fonds Emile Defay' and the Brugmann Foundation for Scientific Research to P.D.

We thank Mr Roger Gheldof for his engineering help as well as Mr T. Renglet who provided the electro-acoustic vibrator and Willy Serniclaes for fruitful advice. Thanks are extended to Nicolas Dumay who acted as speaker. We also thank the editor and two anonymous reviewers for important and helpful comments.

References

- Aaltonen O, Niemi P, Nyrke T, Tuhkanen M. Event-related brain potentials and the perception of a phonetic continuum. *Biol Psychol* 1987;24:197–207.
- Aaltonen O, Tuomainen J, Laine M, Niemi P. Event-related potentials and discrimination of steady-state vowels within phoneme categories. *Scand J Logoped Phonet* 1992;17:107–112.
- Alho K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear* 1995;16:38–51.
- Alho K, Woods DL, Algazi A, Näätänen R. Intermodal selective attention. *Electroenceph clin Neurophysiol* 1992;82:356–368.
- Arnold SA. Objective versus visual detection of auditory brain stem response. *Ear Hear* 1985;6:144–150.
- Bernstein LE, Ponton CW, Auer ETJ. Electrophysiology of unimodal and audiovisual speech perception. *Proceedings of the Auditory–Visual Speech Processing*, Aalborg, Denmark. 2001. p. 50–5.
- Boston JR. Automated interpretation of brainstem auditory evoked potentials: a prototype system. *IEEE Trans Biomed Eng* 1989;36:528–532.
- Burnham D. Language specificity in the development of auditory–visual speech perception. In: Campbell R, Dodd B, Burnham D, editors. *Hearing by eye II: advances in the psychology of speechreading and auditory–visual speech*, New York, NY: Psychology Press, 1998. pp. 29–60.
- Burnham D, Dodd B. Auditory–visual speech perception as a direct process: the McGurk effect in infants and across languages. *Proceedings of the Speechreading by Humans and Machines Conference*. Castéra-Verzudan, France. 1996. p. 103–13.
- Cammann R. Is there a mismatch negativity (MMN) in the visual modality? *Behav Brain Sci* 1990;13:234–235.
- Cheour-Luhtanen M, Alho K, Kujala T, Sainio K, Reinikainen K, Renlund M, Aaltonen O, Eerola O, Näätänen R. Mismatch negativity indicates vowel discrimination in newborns. *Hear Res* 1995;82:53–58.
- Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R. Development of language-specific phoneme representations in the infant brain. *Nat Neurosci* 1998;1:351–353.
- Cheour M, Leppänen PHT, Kraus N. Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clin Neurophysiol* 2000;111:4–16.
- Cole RA, Scott B. Toward a theory of speech perception. *Psychol Rev* 1974;81:348–374.
- Cowan N. *Attention and memory. An integrated framework*, Oxford: Oxford University Press, 1995.
- Cowan N, Winkler I, Teder W, Näätänen R. Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn* 1993;19:909–921.
- Cutting J, Rosner B. Categories and boundaries in speech and music. *Percept Psychophys* 1974;14:591–600.
- Czigler I, Csibra G. Event-related potentials in a visual discrimination task: negative brain waves related to attention and detection. *Psychophysiology* 1990;27:669–676.
- Davis H. Principles of electric response audiometry. *Ann Otol Rhinol Laryngol* 1976;85(suppl 28) pp. 1–96.
- Dehaene-Lambertz G. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 1997;8:919–924.
- Dehaene-Lambertz G. Cerebral specialization for speech and non-speech stimuli in infants. *J Cogn Neurosci* 2000;12:449–460.
- Deltenre P, Mansbach AL. A new descriptor of the dual character of the input–output behaviour of the cochlea, with implications for signal-to-noise ratio estimation of brain-stem auditory potentials evoked by alternating polarity clicks. *Electroenceph clin Neurophysiol* 1993;88:377–388.
- Fodor JA. *The modularity of mind*, Cambridge, MA: MIT Press, 1983.
- Fowler CA. An event approach to the study of speech perception. *J Phonet* 1986;14:2–28.
- Giard MH, Perrin F, Pernier J, Bouchet P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 1990;6:627–639.
- Giard MH, Lavikainen J, Reinikainen K, Perrin F, Bertrand O, Pernier J, Näätänen R. Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: an ERP and dipole-model analysis. *J Cogn Neurosci* 1995;7:133–143.
- Green KP. The use of auditory and visual information during phonetic processing: implications for theories of speech perception. Campbell R, Dodd B, Burnham D, editors. *Hearing by eye II: advances in the psychology of speechreading and auditory–visual speech*, New York, NY: Psychology Press, 1998. pp. 3–25.
- Guthrie D, Buchwald JS. Significance testing of difference potentials. *Psychophysiology* 1991;28:240–244.
- Hardison DM. Bimodal speech perception by native and nonnative speakers of English: factors influencing the McGurk effect. *Lang Learn* 1996;46:3–73.
- Kekoni J, Hämäläinen H, McCloud V, Reinikainen K, Näätänen R. Is the somatosensory N250 related to deviance discrimination or conscious target detection? *Electroenceph clin Neurophysiol* 1996;100:115–125.
- Kraus N, McGee T, Sharma A, Carrell T, Nicol T. Mismatch negativity event-related potential is elicited by speech stimuli. *Ear Hear* 1992;13:158–164.
- Kraus N, McGee T, Carrell TD, Sharma A. Neurophysiologic bases of speech discrimination. *Ear Hear* 1995;16:19–37.
- Kraus N, McGee T, Carrell TD, Zecker SG, Nicol TG, Koch DB. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 1996;273:971–973.
- Kuhl PK. Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech–sound categories. *J Acoust Soc Am* 1981;70:340–349.
- Lang AH, Eerola O, Korpilahti P, Holopainen I, Salo S, Aaltonen O. Practical issues in the clinical application of mismatch negativity. *Ear Hear* 1995;16:117–129.
- Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition* 1985;21:1–36.
- Lieberman AM. When theories of speech meet the real world. *J Psycholing Res* 1998;27:111–122.
- Lieberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. *J Exp Psychol* 1957;54:358–368.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev* 1967;74:431–461.

- Maiste AC, Wiens AS, Hunt MJ, Scherg M, Picton TW. Event-related potentials and the categorical perception of speech sounds. *Ear Hear* 1995;16:68–90.
- Massaro DW. Speech perception by ear and eye: a paradigm for psychological inquiry, Hillsdale, NJ: Lawrence Erlbaum Associates, 1987.
- Massaro DW. Perceiving talking faces. From speech perception to a behavioral principle, Cambridge, MA: MIT Press, 1998.
- Mattingly IG, Liberman AM. Specialized perceiving systems for speech and other biologically significant sounds. In: Edelman GM, Gall WE, Cowan WM, editors. Auditory function: neurobiology bases of hearing, New York, NY: Wiley, 1988. pp. 775–793.
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 1976;264:746–748.
- McPherson DL, Andrews SM. Mismatch negativity to auditory and visual discrepancies to three phonemes. Abstracts of the XV Biennial Symposium of the IERASG. Memphis, TN. June 1997.
- Miller J, Wier C, Pastore R, Kelly W, Dooling R. Discrimination and labelling of noise buzz sequences with varying buzz lead times: an example of categorical perception. *J Acoust Soc Am* 1976;60:410–417.
- Näätänen R. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci* 1990;13:201–288.
- Näätänen R. Attention and brain function, Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.
- Näätänen R. The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear* 1995;16:6–18.
- Näätänen R, Alho K. Mismatch negativity – a unique measure of sensory processing in audition. *Int J Neurosci* 1995;80:317–337.
- Näätänen R, Alho K. Mismatch negativity – the measure for central sound representation accuracy. *Audiol Neuro-Otol* 1997;2:341–353.
- Näätänen R, Lehtokoski A, Lennes M, Cheour-Luhtanen M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 1997;385:432–434.
- Novak GP, Ritter W, Vaughan Jr HG, Wiznitzer ML. Differentiation of negative event-related potentials in an auditory discrimination task. *Electroenceph clin Neurophysiol* 1990;75:255–275.
- Nyman G, Alho K, Laurinen P, Paavilainen P, Radil T, Reinikainen K, Sams M, Näätänen R. Mismatch negativity (MMN) for sequences of auditory and visual stimuli: evidence for a mechanism specific to the auditory modality. *Electroenceph clin Neurophysiol* 1990;77:436–444.
- Paavilainen P, Alho K, Reinikainen K, Sams M, Näätänen R. Right hemisphere dominance of difference mismatch negativities. *Electroenceph clin Neurophysiol* 1991;78:466–479.
- Pause BM, Krauel K. Chemosensory event-related potentials (CSERP) as a key to the psychology of odors. *Int J Psychophysiol* 2000;36:105–122.
- Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept Psychophys* 1973;13:253–260.
- Ponton CW, Don M. The mismatch negativity in cochlear implant users. *Ear Hear* 1995;16:131–146.
- Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical recipes in Pascal. The art of scientific computing, Cambridge, MA: Cambridge University Press, 1990.
- Remez RE, Rubin PhE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychol Rev* 1994;101:129–156.
- Ritter W, Deacon D, Gomes H, Javitt DC, Vaughan Jr HG. The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review. *Ear Hear* 1995;16:52–67.
- Rosenblum LD, Saldaña HM. An audiovisual test of kinematic primitives for visual speech perception. *J Exp Psychol Hum* 1996;22:318–331.
- Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. *Percept Psychophys* 1997;59:347–357.
- Sams M, Aulanko R, Aaltonen O, Näätänen R. Event-related potentials to infrequent changes in synthesized phonetic stimuli. *J Cogn Neurosci* 1990;2:344–357.
- Sams M, Aulanko R, Hämäläinen M, Hari R, Lounassmaa OV, Lu ST, Simola J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 1991;127:141–145.
- Schröger E. On the detection of auditory deviants: a pre-attentive activation model. *Psychophysiology* 1997;34:245–257.
- Sharma A, Kraus N, McGee T, Carrell T, Nicol T. Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroenceph clin Neurophysiol* 1993;88:64–71.
- Stevens KN. On the quantal nature of speech. *J Phonet* 1989;17:3–45.
- Summerfield Q, McGrath M. Detection and resolution of audiovisual incompatibility in the perception of vowels. *Q J Exp Psychol* 1994;36A:51–74.
- Sussman E, Ritter W, Vaughan Jr HG. Predictability of stimulus deviance and the mismatch negativity. *NeuroReport* 1998a;9:4167–4170.
- Sussman E, Ritter W, Vaughan Jr HG. Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Res* 1998b;789:130–138.
- Tales A, Newton P, Troscianko T, Butler S. Mismatch negativity in the visual modality. *NeuroReport* 1999;10:3363–3367.
- Uwer R, von Suchodoletz W. Stability of mismatch negativities in children. *Clin Neurophysiol* 2000;111:42–52.
- Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, Aaltonen O, Raimo I, Alho K, Lang H, Iivonen A, Näätänen R. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn Brain Res* 1999;7:357–359.
- Woods DL, Alho K, Algazi A. Intermodal selective attention. *Electroenceph clin Neurophysiol* 1992;82:341–355.