The whole is more than the sum of its parts - audiovisual processing of phonemes investigated with ERPs

Dörte Hessler[a,b,*], Roel Jonkers[a], Laurie Stowe[a,b], Roelien Bastiaanse[a]

[a] *Center for Language and Cognition Groningen, University of Groningen, The Netherlands*
[b] *Neuroimaging Center, University of Groningen, The Netherlands*

Abstract

In the current ERP study, an active oddball task was carried out, testing pure tones and auditory, visual and audiovisual syllables. For pure tones, an MMN, an N2b, and a P3 were found, confirming traditional findings. Auditory syllables evoked an N2 and a P3. We found that the amplitude of the P3 depended on the distance between standard and deviant. A smaller distance required more attention, which was reflected in a larger amplitude. An analysis of audiovisual material, after correction for visual activity, showed that McGurk type stimuli evoked brain responses that differed from both the standard and the congruent deviants. Finally, we found that congruent audiovisual stimuli elicited an N2 with a shorter latency and a P3 with a smaller amplitude than auditory stimuli. The current ERP study, thus, shows that for audiovisual processing the whole is more than the sum of its parts.

Keywords: speechreading, phonemic dimensions, ERP, MMN, P3, active oddball task

*Corresponding author: D. Hessler, Faculty of Arts, University of Groningen, P.O. Box 716, 9700AS Groningen,The Netherlands, Phone: +31-50-3636596, Fax: +31-50-3636855
Email address: me@doerte.eu (Dörte Hessler)

1. Introduction

Language comprehension involves various processing steps, of which the analysis and identification of phonemes forms the first that is specific for language. Processing of phonemes therefore provides the basis of language comprehension. Often when investigating these early phonemic processes, only auditory processes are considered, while there is also an influence of visual information. The articulatory movements of the speaker, when visible, facilitate comprehension (see e.g. Sumby & Pollack, 1954; Reisberg, McLean & Goldfield, 1987). In the current study, we investigate both auditory and audiovisual processing. One of the central aims of this study is to examine the brain activity related to audiovisual processing of different phonemic contrasts. This will be done by using event-related potential (ERP) measures. We focus not only on the pre-attentive discrimination process, but also on activity related to conscious mismatch detection. Before going into more detail on the research questions and the applied methods, we provide a background on audiovisual speech processing and discuss studies using the ERP paradigm.

1.1. Auditory and audiovisual speech processing

Language processing has been described in terms of different models. One of these is the TRACE model (McClelland & Elman, 1986). In this interactive activation model of speech perception, several levels of processing are assumed: a feature level, a phoneme level and a word level. The three levels are fully interconnected. Units within one level are connected via lateral inhibition. Across levels, excitation takes places both bottom-up and

top-down. In a model like this, distances between phonemes are defined on the basis of the feature values that differentiate them. The phoneme /p/ is, for example, closer to the phoneme /t/ than to the phoneme /k/ as the values on the features 'diffuse' and 'burst' are much more similar for the first two.

However, speech is not only perceived auditorily but also visually (speechreading). Evidence for the importance of multimodal processing comes e.g. from a study by Sumby & Pollack (1954), investigating speech perception in noise. Resistance to noise was much higher when speech was presented audiovisually rather than auditorily. More evidence in favor of multimodality in speech perception was added by the findings of McGurk & MacDonald (1976). Participants were presented with non-matching auditory and visual information and had to report their perception. Instead of answering with the auditory (/pa/) or the visual (/ka/) component of the stimulus, they frequently reported a fusion: /ta/. Information gained through speechreading was combined with the auditory information to form a percept, even though there was no necessity (e.g. due to background noise) or instruction to depend on visual information. This phenomenon is known as the 'McGurk effect'.

Campbell (1988, 1990) extended the TRACE model to incorporate audiovisual perception. The feature level has the same acoustic features as the original model, but was extended to include the visually perceived features 'mouth opening' and 'lip-shape'. All features can inhibit each other, regardless of their input modality. Thus, the activation pattern of 'lip-shape' can inhibit certain values of 'diffuse'.

1.2. Event-related potentials (ERPs)

Speech perception can be investigated online with event-related potentials (ERPs), studying neurophysiological activation patterns. Brain reactions to phonemic distinctions and differences between auditory and audiovisual processing can be investigated with an oddball design. In such a design, a sequence of stimuli is presented to the participants. Within this sequence, one of the stimuli, the 'standard', occurs frequently (e.g. 90%). The stimuli occurring in the remaining instances are called 'deviant'. While the 90-10 distribution of stimuli is the standard distribution, also other ratios have been successfully applied (Deacon, Nousak, Pilotti, Ritter & Yang, 1998). A response to the deviants shows that listeners perceive a difference and can be used to investigate what differences are perceptible during phonemic processing.

The mismatch negativity (MMN) is an ERP-component which is elicited by the automatic recognition of a deviating sound (Näätänen, Gaillard & Mäntysalo, 1978). It peaks between 100 and 200ms after the stimulus onset and is largest at frontal electrodes. It is not only found in experiments with tones, but also with linguistic materials, like phonemes (Aaltonen, Niemi, Nyrke & Tuhkanen, 1987; Sams, Aulanko, Aaltonen & Näätänen, 1990), when presented auditorily. It is often claimed that the MMN is specific for auditory input. But there is also evidence for a visual counterpart of the MMN: a negativity in the N2 time window. However, this negativity has a more posterior scalp distribution, suggesting that the effects are generated by different areas in the brain (Pazo-Alvarez, Cadaveira & Amenedo, 2003).

In active oddball tasks, participants are requested to attend to the stim-

uli. When processing an attended deviant stimulus, the MMN is followed by another negativity, the N2b (Näätänen, Simpson & Loveless, 1982). This component is elicited by conscious discrimination tasks and shows an overlapping distribution with the MMN. The N2b is sensitive to auditory and visual deviants, if they are attended. The peak latency is around 200-250ms. The N2b also differs from the MMN in the distribution: the MMN is largest at the frontal electrodes while N2b is largest centrally (Novak, Ritter & Vaughan Jr., 1992). Often the MMN and the N2b are referred to together as the N2 (Luck, 2005).

The P3 follows the N2b with a latency of 300 to 600ms (Courchesne, Hillyard & Galambos, 1975). It has a broad distribution, but is largest at the parietal electrodes. The P3, just like the N2b, is only elicited when the stimuli are attended and occurs with both visual and auditory deviants. The amplitude of the P3 is related to the stimulus probability (Duncan-Johnson & Donchin, 1977), the resources allocated to the task (Isreal, Chesney, Wickens & Donchin, 1980) and the uncertainty (equivocation) of the participants (Johnson, 1984, 1986). The contribution of probability (P), equivocation (E), and resource allocation (R) to the overall P3 amplitude was formalized by Johnson (1984, 1986) in the following manner:

$$\text{P3 amplitude} = E \times (P + R)$$

None of the components discussed above is specific to language processing. Nonetheless, they all have been found to react to phonemic differences as well. Thus, the active oddball technique is an appropriate tool to assess phoneme processing.

1.2.1. Phoneme processing in ERP studies

Phonological processing has been addressed with oddball studies. Lawson & Gaillard (1981) showed with an active oddball design that the peak latency and amplitude of mismatch responses were influenced by the number of phonetic dimensions differing between standard and deviant. The N2 was found to be a good indication of phonetic distance: the larger the distance, the shorter the latency and the higher the amplitude of the N2. In this study, distinctions of different sizes (different number of dimensions) were investigated, but contrasts within one dimension were not looked at.

Processing differences of voice onset time and place of articulation have also been analyzed in oddball designs. Several studies reported a relationship between the amplitude and/or latency of the MMN and the magnitude of the difference between standard and deviant on a 'place of articulation' continuum, but no effect of categorical perception has been found. The MMN changed continuously between different exemplars of the same phoneme and did not show a larger effect at a phoneme border (Sams et al., 1990; Kraus, McGee, Sharma, Carrell & Nicol, 1992; Sharma, Kraus, McGee, Carrell & Nicol, 1993; Maiste, Wiens, Hunt, Scherg & Picton, 1995). Sharma & Dorman (1999) were, however, able to find a category effect for a 'voice onset time' continuum.

Investigations of audiovisual processing have concentrated on the McGurk effect and were done with passive oddball designs. Möttönen, Krause, Tiippana & Sams (2002) conducted a magnetic encephalographic (MEG) study with congruent and incongruent audiovisual stimuli. The incongruent stimuli differed only in the visual part from the standard. In MEG, this kind of

paradigm evokes a so-called 'mismatch field' (MMF), which is comparable to the MMN in ERP studies. Both congruent and incongruent deviants elicited an MMF which had a larger amplitude for congruent than for incongruent deviants. The latency was shorter for the audiovisual incongruent deviants than for purely visual stimuli without auditory input. This suggests that the interaction between auditory and visual information accelerates the detection of deviants.

Studies investigating the McGurk effect with the ERP paradigm aimed to prove that the visual information alters auditory perception (e.g. Colin, Radeau, Soquet, Demolin, Colin & Deltenre, 2002; Colin, Radeau, Soquet & Deltenre, 2004; Saint-Amour, De Sanctis, Molholm, Ritter & Foxe, 2007). Colin et al. (2002, 2004) found that an MMN was evoked for both auditory and McGurk stimuli, although the McGurk stimuli were auditorily identical to the standard. As the visual mismatch within the McGurk condition was not expected to elicit an MMN, the authors concluded that the auditory perception of the participants was altered and therefore perceived as deviant although the auditory part of the stimulus was not different from the standard.

Similar results were found by Saint-Amour et al. (2007) in an oddball study with visual and audiovisual stimuli. The deviant stimuli were the visual syllable /va/ and the incongruent audiovisual syllable /ba/[va] (auditory /ba/ dubbed on visual syllable /va/), which lead to the perception of /va/. The standard stimulus was the syllable /ba/ (either visual or audiovisual congruent). There was no mismatch response for the visual stimuli, but ongoing visual activity for standards and deviants. The visual activity was

subtracted from the audiovisual activity in order to avoid an overlay of the response to the visual processing in the McGurk condition. A typical MMN was found at Fz. However, since there was no direct comparison with the auditory MMN, it is not clear whether the effects are in fact identical.

Most of the studies reported above were carried out with passive oddball tasks and focused on automatic processes. In the current study, we will extend the paradigm to an active oddball design, in order to investigate also conscious processing. This will provide valuable additional information, such as the response times and the accuracy rates per stimulus type. Furthermore, false alarms (for the standards) and misses (for the deviants) can be excluded from the analysis.

With the design adapted in the way described above, we aim to address the following issues:

(1) We will investigate whether the amplitude of any of the ERP responses to a phonemic deviant depends on the size of the mismatch, as has been reported for the processing of tones. The difference between standard and deviant will be within the dimension 'place of articulation', regarding the features 'diffuse', 'acute' and 'burst', as they were defined in the TRACE model (McClelland & Elman, 1986).

(2) Furthermore, we will study the integration of auditory and visual information. With the use of both congruent and McGurk type deviants, we will investigate whether the activity related to the integration of incongruent audiovisual information is similar to the activity for audiovisual integration of congruent stimuli.

(3) Finally, we want to find out whether brain responses represent the in-

tegration of auditory and visual information by looking at audiovisual processing of congruent information. We will evaluate whether the response to audiovisual processing is more than a mere addition of the brain waves to auditory and visual processing.

2. Methods

In order to address the issues stated above, an active oddball experiment was carried out in four different variants: 'pure tones', 'auditory syllables', 'visual syllables' (videos of articulatory movements), and 'audiovisual syllables'. Participants were asked to identify infrequent deviant stimuli in a series of repeating standard stimuli.

2.1. Participants

Thirteen native speakers of Dutch (nine female) participated in this study after giving their informed consent. None of them reported neurological, language or hearing disorders. Vision was normal or corrected to normal in all individuals. All participants were right-handed. The mean age was 59 years (range 45-69). The mean age in this study deviates from that of other reported studies due to the fact that the current study was part of a larger project investigating auditory and audiovisual processing in aphasia. Therefore the age of the participants is matched with those of the aphasic participants in the larger project.

2.2. Materials

In the first three sub-experiments a sequence of standard stimuli and two deviant stimuli was presented. There were 800 repetitions of the standard

stimulus (80 % of all stimuli) and 100 repetitions of each deviant (each 10 % of all stimuli). In the audiovisual sub-experiment, a McGurk type deviant was added. Therefore, there were three deviants, each of which occurred 100 times (6.66 % of all stimuli). The standard was repeated 1200 times, forming 80 % of stimuli. This way the proportion of standards and deviants (overall, not per deviant), as well as the absolute number of items for each deviant type, were kept constant across sub-experiments.

The stimuli in the 'tones' sub-experiment were generated with the computer program Praat (Boersma & Weenink, 2009). The standard stimulus was a pure tone of 1000Hz, the deviant stimuli were pure tones of 1050Hz (near deviant) and 1200Hz (distant deviant). They were presented auditorily while displaying a white screen with a fixation cross in the middle to minimize eye movements during trials.

The stimuli in the remaining three sub-experiments were the standard /pa/ and the deviants /ta/ (near) and /ka/ (distant) as these syllables differ in only one dimension ('place of articulation') and are the ones involved in the McGurk effect. As described above, /pa/ and /ta/ are distinguished by the features 'acute' and 'burst', while the distinction between /pa/ and /ka/ is based on differences in 'acute', 'burst' and 'diffuse'. Overall, the distance between the latter two is larger. In the audiovisual sub-experiment, an audiovisual incongruent syllable, eliciting the McGurk effect (auditory /pa/ dubbed on visual /ka/) was added.

The syllables were spoken by a male native speaker of Dutch, who was video-recorded in a quiet room with daylight. Additionally, a light diffuser was used to avoid shading on the face for optimal visual information. The

recorded image included the lower part of the speaker's face (from the lower part of the nose), the neck and the shoulders. For recording, a video camera and separate cardioid microphone were used. The video was then digitized into avi-files at a sampling rate of 48 kHz with 32-bit-stereo quantization. All stimuli were then edited with Adobe Premiere to form video files with a duration of 800ms each. As recording was done with 25 frames per second (i.e. the duration of one frame is 40ms), each file consisted of 20 frames. The video showed the speaker in rest (with a closed mouth) for 6 frames (240ms) in the beginning of each video (baseline for movement). The initial preparatory movements of the mouth lead the sound onset by 200ms, on average (range: 180-220ms).

In the 'auditory syllables' sub-experiment, the stimuli were presented with a white screen with a fixation cross replacing the speakers face. In the 'visual only' sub-experiment, the videos were played without sound, showing only the articulatory movements of the speaker. In the 'audiovisual' sub-experiment, both sound and articulatory movements were presented. The sound and articulatory movements were congruent for the standard and two of the deviants and were incongruent for the McGurk deviant.

2.3. Procedure

The experiment was set up as an active oddball task. Standard and deviant stimuli were presented in a semi-randomized order: each deviant was preceded by at least three and maximally five standards. Stimuli were presented with a stimulus onset asynchrony of 1500ms. Participants were instructed to pay close attention and press a button whenever they detected a deviating syllable. They were told that the first stimulus was a stan-

dard which would occur frequently. In a short practice trial, the procedure was illustrated. Response times and the number of correct detections were recorded. Furthermore, it was automatically recorded if the onset of the video was delayed, which occurred rarely, when the presentation program could not access the rather large video files in time.

All sub-experiments were split into four blocks to ensure that continuous recording time did not exceed 10 minutes, because it has been shown that the MMN decreases due to habituation after 10 minutes (McGee, King, Tremblay, Nicol, Cunningham & Kraus, 2001). Also, the participants needed to pay attention to the stimuli, which made regular breaks necessary.

Testing was carried out on two different days. Two blocks of each sub-experiment were presented per day. All participants started with the two blocks of the 'tones' sub-experiment. The order of the other sub-experiments was balanced between participants and recording days.

The volume of the stimuli was kept constant for all participants at 65dB. The screen refresh rate and the triggers sent to the EEG for segmentation were measured and compared with an oscilloscope. There was alignment between the refresh rate and the triggers, ensuring synchrony of the video onset with the trigger.

2.4. EEG recording and analysis

A 64-electrode elastic cap (Electro-Cap International) with tin electrodes was used to record the EEG. Reference electrodes were placed on both mastoid bones. A ground electrode, placed on the sternum, served as common reference. Bipolar horizontal and vertical electrooculograms (EOG) were recorded and used to correct for eye movements.

The mean impedance over all participants of the different electrodes varied from 2KΩ to 6.5KΩ. The impedance of individual electrodes per participant was kept below 10KΩ, with a few exceptions. The highest impedance of an electrode included in the analysis was 17KΩ, for one electrode of one participant. EEG and EOG signals were recorded with Brain Vision Recorder (Brain Products GmbH) and sampled at 250Hz.

The analysis of the EEG data was done with Brain Vision Analyzer 1.05 (Brain Products GmbH). In a first step, both recordings of one participant were combined to one dataset. The reference during recording was an average of all electrodes; the data were re-referenced off-line with the two mastoid electrodes as reference [1]. The data were furthermore filtered with a low cut-off point of 0.1Hz and a high cut-off point of 50Hz. Ocular correction for blinks was carried out using the Gratton-Coles method. Semi-automatic artifact rejection was applied with the following automatic parameters: maximum voltage change on a single step was 50μV, maximum difference in values within 200ms was 200μV and the maximal amplitude was 200μV. All segments were inspected and additional irregularities were marked as artifacts. So as not to include more standards than deviants in the statistical comparison, only standards directly preceding a deviant were included in the analysis. These were also the standards which were most prototypical,

[1] The mastoids rather than an average reference were chosen for the following reason: The mastoids or earlobes serve as reference in most studies investigating phonemic processing with oddball tasks. To warrant the best possible comparability, we also decided against an average reference. Furthermore we were also concerned about the amount of electrodes available for averaging. A too low density would lead to unreliable results.

as they are always preceded by at least two other standards. Trials with a delay in video presentation were excluded from the analysis, as were trials where the participants made errors (i.e. missed a deviant or responded to a standard). For all analyses, a baseline was set from 200ms before stimulus onset until the onset itself. Individual averages were calculated and served as a basis for the grand averages. An additional high cut-off filter of 10Hz was applied to the grand averages for use in figures only.

In the 'tones' and 'auditory syllables' sub-experiments, the onset was the beginning of the sound. The time windows investigated were from 120 to 160ms after sound onset for the MMN, between 200 and 240ms for the N2b and from 360 to 400ms. These time intervals were chosen based on the literature. For the 'visual only' sub-experiments the same intervals were used, but taking the onset of visual differences as starting point. In the 'audiovisual' sub-experiment, the target onset (0) is the auditory onset which lags behind the visual onset by 200ms. The intervals were chosen with reference to the auditory onset and are equivalent to those reported above. With regard to the visual onset they are between 320-360, 400-440 and 560-600ms.

For all analyses, nine regions of interest (ROIs) were defined to evaluate for scalp distribution (frontality and laterality, each with three levels). Each ROI consisted of two electrodes[2] for which sufficient data were available after

[2]Two electrodes per region of interest were selected because in midline regions maximally three electrodes were available, with FPz being of rather poor quality, for a number of subjects. Therefore, in the other regions also only two electrodes were chosen. These were the ones with best quality and, if possible, most 'prototypicality' for the region, for example the most central for the central ROI (C3 rather than FC3 or CP3).

removing stimuli with a video onset delay, incorrect answers and artifacts. Electrodes in the left and right ROIs were mirrored (see Figure 1).

[Place Figure 1 about here]

For each time window, three-way repeated measures ANOVAs (3x3x3) were carried out with the factors frontality (frontal, central, occipital), laterality (left, midline, right) and stimulus type (standard, deviant1, deviant2). The scalp distribution factors were only of interest if they interact with the factor stimulus type, so significant effects limited to these two factors will not be reported. Whenever a main effect of stimulus type was found, pairwise comparisons were carried out to determine which of the three types led to the effect. While the test statistics and the significance value were calculated based on degrees of freedom corrected for sphericity (Greenhouse-Geisser correction), the uncorrected degrees are provided in this paper, for the sake of readability.

The behavioral results were analyzed with regard to the number of correct answers per sub-experiment and stimulus type and the response times when detecting a deviant. Repeated measures ANOVAs with the factor 'stimulus type' (standard, deviant1, deviant2, McGurk) were carried out with regard to accuracy. Posthoc pairwise comparisons were carried out when the ANOVA yielded significant results. For the response times, repeated measures ANOVAs with three levels of 'stimulus type' (deviant1, deviant2, McGurk) were carried out, which were followed-up by pairwise comparisons if significant.

3. Results

3.1. Behavioral results

Both the accuracy and the response times were recorded for the detection of the different deviants. For the standard, only accuracy can be reported, as the correct response was to not push a button. Participants showed hardly any false alarms, but missed some of the deviants. An overview of the behavioral results is given in Table 1.

Table 1: Percentage correct and mean response times for the different sub-experiments and stimuli. For McGurk stimuli, button pushes were counted as correct answers.

Condition	Standard (/pa/ or 1000Hz)	Deviant 1 (/ka/ or 1050Hz)		Deviant 2 (/ta/ or 1200Hz)		McGurk deviant (/pa/[ka])	
	% correct	% correct	RT	% correct	RT	% correct	RT
Tones	99.9%	94.3%	530ms	99.3%	450ms	-	-
Auditory Syll.	99.9%	74.7%	912ms	80.6%	909ms	-	-
Visual Syll.	99.9%	93.0%	701ms	96.9%	669ms	-	-
Audiovisual Syll.	99.9%	92.1%	784ms	96.1%	743ms	86.8%	782ms

The accuracy of the reactions of the participants differed per stimulus type in the auditory ($F(2,24)=15.312$, $p<0.001$), the visual ($F(2,24)=5.756$, $p<0.05$) and the audiovisual sub-experiments ($F(3,36)=11.998$, $p<0.01$), but not for the tones sub-experiment ($F(2,24)=1.605$, $p=0.229$). Pairwise comparisons revealed that there are more errors for both kinds of deviant than for the standards in the auditory sub-experiment ($p<0.01$, for both comparisons: standard – deviant /ka/ and standard – deviant /pa/). The accuracy did not differ significantly between deviant types ($p=0.074$). In the visual sub-experiment, the pairwise comparisons showed a different pattern: the accuracy for the standard is higher than for the deviant /ka/ ($p<0.01$), but

does not differ significantly from the deviant /ta/ ($p=0.067$). The number of correct responses for the deviant /ka/ is also significantly lower than for the deviant /ta/ ($p<0.01$). In the audiovisual sub-experiment, all pairwise comparisons are significant on at least the 0.05 level. Accuracy is the highest for standards, followed by the deviant /ta/, then the deviant /ka/ and is lowest for the McGurk stimuli (when button pushes are counted as correct answer).

Response times were compared for the deviants of all four conditions. We assume that response times reflect the certainty of the decision. The more certain participants are about their decision the faster they will respond (cf. Morin & Forrin, 1963). Certainty (equivocation) is one of the measures that influences the amplitude of the P3 component. The deviants differed significantly from each other with regard to the response times in the tones sub-experiment ($F(1,12)=32.414$, $p<0.001$) and the visual sub-experiment ($F(1,12)=16.951$, $p<0.001$). The factor stimulus type also influenced the response time in the audiovisual sub-experiment ($F(2,24)=13.731$, $p<0.01$): reactions to deviant 2 (/ta/) were faster than to both deviant 1 (/ka/, $p<0.01$) and the McGurk stimulus ($p<0.01$). The latter two stimulus types did not differ from each other ($p=0.811$). In the auditory sub-experiment, there was no difference in the response times to the deviant stimuli ($F(1,12)=1.066$, $p=0.322$).

3.2. Pure tones

Figure 2 depicts the activity recorded in the 'tones' sub-experiment. In the first time window both deviants evoked a more negative response than the standard at the frontal and central electrodes. The effect is strongest at

frontal electrodes (see left panel of Figure 2. In the second time window, both deviants elicited a more negative response than the standard at the central electrode (see middle panel). In the third time window, both deviants show a positivity at posterior electrodes, which is stronger for the more distant deviant.

[Place Figure 2 about here]

MMN (120 to 160ms): A significant effect of stimulus type was found ($F(2,24)=25.674$, $p<0.001$). Pairwise analyses revealed that both deviants differed from the standard ($p<0.001$, for both comparisons), but not from each other ($p=0.126$). Furthermore an interaction of stimulus type and frontality was found ($F(4,48)=11.769$, $p<0.001$): the effect is the strongest for frontal electrodes.

N2b (200 to 240ms): In this time range, no main effect of stimulus type was found ($F(2,24)=0.833$, $p=0.445$). There was, however, an interaction between stimulus type and frontality ($F(4,48)=8.176$, $p<0.01$). This interaction implied an effect of stimulus type limited to central electrodes. When analyzing the central electrodes in a two-factor repeated measures analysis (stimulus type by laterality), a main effect for stimulus type was found ($F(2,24)=4.913$, $p<0.05$): both deviants differed from the standard ($p<0.05$, for both comparisons), but not from each other ($p=.242$).

P3 (360 to 400ms): A significant main effect of stimulus type ($F(2,24)=16.072$, $p<0.001$) was found: there were significant differences between the standard and the more distant deviant (1200Hz, $p<0.001$) and between both deviants ($p<0.01$) Furthermore, there was a three-way interaction between laterality, frontality and stimulus type ($F(8,96)=3.658$, $p<0.05$).

While the effect for stimulus type appeared to be largest in the midline occipital region, there is clearly also an positive upswing visible at central electrodes which partially overlapped with the N2b negativity. In the occipital region, there was a significant main effect for stimulus type ($F(2,24)=16.097$, $p<0.001$). Also, all three pairwise comparisons were significant ($p<0.01$, for all three comparisons), indicating differences between the standard and both deviants and between the deviants.

For completeness, also the other time intervals were analyzed. We only looked for main effects of stimulus type and can report that the effect described for the MMN time-window, was already recordable from 80ms on and lasted until 200ms. At 240-280ms we also found a negativity for the 1050Hz deviant. The positivity for the 1200Hz deviant (P3 time window) was recordable from 320ms onwards.

3.3. Auditory syllables

Figure 3 depicts the ERP activity for the auditory syllables. In the first time window, no clear effect can be seen. In the second time window, there is a negativity for both deviants in the frontal and central electrodes (left and middle panels), which starts around 250ms for the frontal and around 200ms for the central electrode. In the third time window, a positivity for both deviants can be seen at all three locations, which appears larger for deviant /ta/ than deviant /ka/ at frontal and central electrodes.

[Place Figure 3 about here]

MMN (120 to160ms): No main effect of stimulus type was found ($F(2,24)=0.730$, $p=0.474$).

N2b (200 to 240ms): The brain activity depended on the stimulus type ($F(2,24)=9.680$, $p<0.01$). There was also an interaction of stimulus type and frontality ($F(4,48)=4.124$, $p<0.05$): the effect is the strongest for central electrodes. Posthoc pairwise comparisons of the main effect revealed that both deviants differed from the standard (standard - /ka/: $p<0.05$; standard - /ta/: $p<0.01$), but not from each other ($p=0.161$).

P3 (360 to 400ms): We found a main effect of stimulus type ($F(2,24)=34.716$, $p<0.001$): both deviants differed from the standard (standard - /ka/: $p<0.001$; standard - /ta/: $p<0.001$). The effect was larger for the deviant /ta/ than the deviant /ka/ ($p<0.01$). Furthermore, we found an interaction of stimulus type and frontality ($F(4,48)=4.313$, $p<0.05$): the effect is the strongest in frontal electrodes.

Analysis of the remaining time windows revealed that for the deviant /ta/ there was a negativity between 80 and 120ms. There was no significant main effect between the MMN and N2b time windows. The N2b effect for deviant /ta/ was prolonged until 280ms. The reported P3 effect was already detectable at 320-360ms for both deviants.

3.4. Visual syllables

Figure 4 depicts the activity related to the visual syllables. Throughout the time windows, both deviants elicit far more positive waveforms than the standard. Furthermore, the deviants also differ from each other: in the second and third time window, the deviant /ta/ elicits a larger positivity than the deviant /ka/. The figure also illustrates how early the positivity starts and how long it lasts. This underlines the major effect that deviancy had in this sub-experiment. Note that unlike the previous sub-experiments, there is

activity in the baseline period that is related to visual input. This is due to the fact that the measurement was time-locked to the onset of articulatory movement. The video, however, started earlier, showing the speaker in rest for 240ms before articulatory movements set in, evoking a response to the visual input. Since the movement is part of the same visual event, it did not evoke such clear early components as the initial part of the event. There also seem to be differences between standard and deviant stimuli in this time-period. This is confirmed by statistical testing (main effect of stimulus type: $F(2,24)=7.490$, $p<0.01$; pairwise post-hoc testing: both deviants are slightly more positive than the standard: $p<0.01$ for /ka/ and $p<0.05$ for /ta/). There is no evident reason for this difference. Visual inspection of the results of Saint-Amour et al. (2007) showed a similar effect during the baseline period. They, however, do not address the issue. One reason for the difference between standards and deviants could be small differences in the resting phase that were not noticed during video inspection. The visual condition, however, was not directly used to answer the research questions, but rather served as a control condition, necessary to correct the audiovisual condition for visual activity. The phenomenon described above once more illustrates the necessity of correcting for this early starting and ongoing visual activity.

[Place Figure 4 about here]

MMN (120 to 160ms): In the MMN-time window, a main effect of stimulus type was found ($F(2,24)=8.119$, $p<0.01$): both deviants differed from the standard (deviant /ka/: $p<0.01$, deviant /ta/: $p<0.05$). Furthermore, there was a three-way interaction between frontality, laterality and

stimulus type ($F(8,96)=6.605$, $p<0.01$), indicating that this difference was largest at left frontal electrodes.

N2b (200 to 240ms): A main effect of stimulus type ($F(2,24)=27.43$, $p<0.001$) and an interaction of stimulus type with laterality ($F(4,48)=5.735$, $p<0.01$) as well as the three-way interaction between laterality, frontality and stimulus type ($F(8,96)=4.664$, $p<0.01$) were found. The posthoc pairwise comparison of the main effect revealed that both deviants differed from the standard (standard vs. /ka/: $p<0.01$; standard vs. /ta/: $p<0.001$). They also differed from each other ($p<0.05$). The effect appeared largest at midline electrodes.

P3 (360 to 400ms): There was a significant main effect for stimulus type ($F(2,24)=53.293$, $p<0.001$). The interaction between stimulus type and laterality ($F(4,48)=12.533$, $p<0.001$) and the three-way interaction ($F(8,96)=6.111$, $p<0.01$) were also significant. They showed that the effect was strongest occipitally around the midline. The posthoc pairwise comparison of stimulus type revealed that both deviants differed from the standard ($p<0.001$ for both comparisons). The two deviants also differed significantly from each other ($p<0.01$). An analysis of the other time windows showed that the positivity recorded in the MMN, N2b and P3 time windows were detectable from 80ms onwards (both deviants more positive than the standard). The difference between the two deviants was first recorded in the N2b time window, but then carried on until the P3 time window (deviant /ta/ more positive than deviant /ka/ and both more positive than the standard).

3.5. Audiovisual syllables

As can be seen from the visual sub-experiment, visual mismatches have an effect in an active oddball task. The visual input leads the auditory by 200ms. Therefore, in the time window where the visual deviance elicits a P3, the auditory deviance elicits a negativity. This had to be taken into account in the analysis of this sub-experiment. Therefore, the activity of the visual stimuli was subtracted from the audiovisual stimuli (hereafter called 'corrected audiovisual'), to remove the visual mismatch effect. For the McGurk deviant (visual /ka/, auditory /pa/), the activity of the visual deviant /ka/ was subtracted (cf. Saint-Amour et al., 2007). The remaining activity should then be due to the auditory part /pa/ or additional audiovisual integration activity, the saliency and the ease of deviant detection. The reported intervals are based on the onset of the auditory difference. For completeness, however, the original, uncorrected waves of the audiovisual condition are depicted in Figure 5.

[Place Figure 5 about here]

3.5.1. McGurk stimuli

For the evaluation of the McGurk effect, the comparisons were carried out with three-way repeated measures ANOVAs with the factors frontality (3 levels), laterality (3 levels), and stimulus type (4 levels: standard, deviant /ka/, deviant /ta/, and McGurk deviant).

6, the waves of all four stimuli are depicted. The McGurk stimulus showed a more negative waveform than the standard at all three electrode locations, which started in the first and the second time window at occipital electrodes.

In the third time window, this negativity was also found at the other electrode locations. Furthermore, the McGurk stimulus was then also more negative than the other two deviants, while they appeared quite equal in the earlier time windows.

[Place Figure 6 about here]

MMN (120 to 160ms): There was no main effect of stimulus type ($F(3,36)=1.469$, $p=0.25$), but an interaction of stimulus type and frontality ($F(6,72)=4.677$, $p<0.05$). Separate two-way ANOVAs for each instance of frontality revealed a significant main effect of stimulus type at the occipital electrodes ($F(3,36)=4.191$, $p<0.05$), a trend at frontal electrodes ($F(3,36)=2.817$, $p=0.065$) and no significant difference at central electrodes ($F(3,36)=1.838$, $p=0.175$). Pairwise comparisons showed that the McGurk stimulus differed from the standard at the occipital electrodes ($p<0.01$), despite being auditorily identical. Moreover, the congruent deviant /ka/ also elicited a more negative response than the standard ($p<0.05$).

N2b (200 to 240ms): There was a main effect of stimulus type ($F(3,36)=6.908$, $p<0.01$): the response evoked by the McGurk stimulus differed from the responses evoked by the standard ($p<0.01$) or deviant /ta/ ($p<0.05$). Also, the deviant /ka/ showed a significantly more negative reaction than the standard ($p<0.01$). An interaction between stimulus type and frontality ($F(6,72)=4.303$, $p<0.05$) indicated that effect was strongest in the central and the occipital regions and less so in the frontal region. Also, an interaction between stimulus type and laterality was found ($F(6,72)=3.958$, $p<0.01$): the difference between McGurk and standard stimuli was largest around the midline.

P3 (360 to 400ms): Between 360 and 400ms, we found a main effect of stimulus type ($F(3,36)=5.878$, $p<0.01$): the McGurk stimulus differed from each of the other three stimuli: the standard ($p<0.01$), deviant /ka/ ($p<0.01$) and deviant /ta/ ($p<0.01$). The two deviants differed neither from the standard nor from each other (deviant /ka/ - standard: $p=0.536$, deviant /ta/ - standard: 0.592, /ka/ - /ta/: 0.782). A significant interaction between stimulus type and frontality ($F(6,72)=8.485$, $p<0.001$) indicated that the difference between the McGurk stimulus and the other stimuli is located in the central and occipital regions. Additionaly to the effects reported above we found that the McGurk deviant elicited a more negative response than the standard from 40-80ms, a more negative response than all other item types from 24-280ms, a more negative response than the standard and the deviant /ka/ from 280-320ms and a more negative response than the deviant /ka/ from 320-360ms. The only other main effect of stimulus type found was between 160 and 200ms, where deviant /ka/ evoked a more negative response than the standard stimulus.

3.5.2. Comparison of auditory and audiovisual processing

The main interest in this sub-experiment was to investigate whether activation can be found that is additional to the activation from the separate auditory and visual inputs. Therefore, a further subtraction was applied: For both the auditory and the corrected audiovisual activity the difference between deviant and standard was calculated. These difference waves made it possible to compare the auditory part of the audiovisual deviance response to the pure auditory deviance response.

Figure 7 depicts the difference waves for both presentation modalities for

each deviant. For the deviant /ka/, it can be seen that the negativity starts earlier for the 'corrected audiovisual' stimuli than for the auditory and that the positivity is larger for the auditory difference. For the deviant /ta/ too, the 'corrected audiovisual' negativity has an earlier onset than the auditory negativity and the positivity is larger in the auditory modality than in the 'corrected audiovisual' modality. This time-shift in the negativity cannot easily be directly caught when analyzing individual time windows, but only becomes apparent in visual inspection. The two presentation modalities were statistically compared in order to support the findings from visual inspection.

[Place Figure 7 about here]

The time windows analyzed in this sub-experiment refer to the onset of the auditory difference, which is 200ms later than the visual difference. A four-way ANOVA with the factors presentation modality (2: corrected audiovisual and auditory), frontality (3), laterality (3), and deviant (2: deviant /ta/ and deviant /ka/) was carried out to investigate whether there were differences in the brain responses.

120 to 160ms: No main effect for modality ('corrected audiovisual' versus auditory) was found ($F(1,12)=0.428$, $p=0.525$). There was, however, an interaction of modality and frontality ($F(2,24)=12.229$, $p<0.001$). Separate three-way ANOVAs for each level of frontality revealed that there was no main effect of modality at either frontal ($F(1,12)=1.709$, $p=0.216$), central ($F(1,12)=1.027$, $p=0.331$) or occipital electrodes ($F(1,12)=3.812$, $p=0.075$).

200 to 240ms: No main effect of modality was found ($F(1,12)<1$). There were, however, interactions between modality and frontality ($F(2,24)=7.559$,

$p<0.01$) and between modality and deviant ($F(1,12)=10.94$, $p<0.01$) When looking only at the deviant /ka/, no main effect of modality was found ($F(1,12)=1.63$, $p=0.226$), but again the interaction between modality and frontality emerged ($F(2,24)=7.458$, $p<0.01$). For the other deviant, /ta/, no main effect of modality ($F(1,12)=0.856$, $p=0.373$) and no significant interaction of modality with frontality were found ($F(2,24)=2.927$, $p=0.094$). Three-way ANOVAs for each level of frontality showed that no main effect of modality could be found for any level (frontal: $F(1,12)=3.123$, $p=0.103$; central: $F(1,21)=0.072$, $p=0.793$; occipital: $F(1,12)=2.507$, $p=0.139$).

360 to 400s: In this time window, there was a significant main effect of modality ($F(1,12)=35.38$, $p<0.001$). Furthermore, we found a significant interaction with deviant type ($F(1,12)=5.277$, $p<0.05$), indicating that the difference between the presentation modalities was larger for the deviant /ta/ than for the deviant /ka/.

Additionally to the effects for the major time windows, we found a main effect of modality between 160 and 200ms and between 320 and 360ms (the audiotory stimuli evvoked a more positive response).

3.6. Summary of results

Several analyses have been carried out to address the issues raised above. This summary provides the results with regard to each sub-experiment.

Tones: For the pure tones, the accuracy was equally high for the standard and both deviants. Participants reacted more quickly to the more distant deviant than to the less distant deviant. In the ERP, we found an MMN, which was strongest at frontal electrodes, an N2b (at central electrodes) and a P3 (strongest at occipital electrodes) for both deviants.

Auditory syllables: For the auditory syllables, participants showed a higher accuracy for the standards than for either type of deviant. Responses to the deviants differed neither in accuracy nor in response time from each other. No ERP influence of stimulus type was found in the MMN time-window. In the period between 200 and 240ms, a significant negativity was found for the deviants (especially at central electrodes). Also, a P3 was found, which was larger for the less distant deviant /ta/ than for the deviant /ka/.

Visual syllables: In the visual sub-experiment, the accuracy was higher for the standards and deviant /ta/ than for the deviant /ka/. Also, the response time was shorter for deviant /ta/ than for deviant /ka/. The brain response to the deviant syllables was more positive than the response to the standard syllables throughout the different time-windows. The effect had a broad scalp distribution, but was largest occipitally. The deviant /ta/ evoked a more positive response than the deviant /ka/ in the time slots between 200 and 240 and between 360 and 400ms.

McGurk syllables: In the audiovisual sub-experiment, the accuracy was highest for the standards, followed by the deviant /ta/, the deviant /ka/ and was lowest for the McGurk deviant when the correct answer is regarded as pushing a button. Also, response times were shortest for deviant /ta/. Deviant /ka/ and the McGurk deviant evoked equally fast responses. Responses to the incongruent McGurk syllables were significantly more negative than to the congruent standards at occipital electrodes in all three time-windows. After correction for visual activity, the McGurk stimulus elicited also more negative responses than the deviants which were deviant in both the visual

and auditory modalities (/ka/ and /ta/) between 360 and 400ms.

Audiovisual versus auditory syllables: The comparison of the brain responses between the auditory part of the audiovisual syllables (after correcting for the deviance response to visual syllables) and the pure auditory syllables revealed that there was a difference between the two presentation modalities, specifically in the latest time-window (360-400ms), with a larger positivity for the auditory only syllables. Also the scalp distribution of the positivity differs between both modalities. While the positivity is limited to frontal electrodes for the corrected audiovisual syllables, it can be found across the scalp for the auditory syllables.

4. Discussion

In this section, we will discuss the results reported above. First, we will discuss the results of the 'tones' sub-experiment, which served as a control for the interpretation of the other results. After that we will address three issues: (1) the degree of deviance in phonemic contrasts, (2) the effect of audiovisual integration on phonemic processing and (3) the brain correlates of the McGurk effect.

4.1. Processing of pure tones

For the pure tones, the participants had no problems detecting both deviants. The more distant deviant (1200Hz) was however detected faster than the less distant one (1050Hz). This decrease in reaction time might reflect a higher certainty about the decision, caused by the larger physical difference.

The ERP findings in the 'tones' sub-experiment resemble the classical findings of for example Sams, Paavilainen, Alho & Näätänen (1985) for active

oddball designs, which also determined the choice of time windows. The effects in the current study are not necessarily limited to the time windows used in the analysis, but these were considered most prototypical. In order to compare the measurements in the other sub-experiments to those taken for the 'tones', the choice of time windows was consistent throughout all sub-experiments. For the 'tones', the three components which were expected were found: the MMN, especially at frontal electrodes, the N2b at central electrodes and a large P3 at occipital electrodes, which was more pronounced for the more distant deviant. Participants showed faster responses indicating a higher certainty for the more distant deviant. The larger P3 amplitude for this deviant can thus be explained by the higher certainty (cf. Johnson, 1984, 1986). The factor probability, which Johnson also mentions as affecting the amplitude of the P3, is equal for both deviants in the current set-up. The third factor 'resource allocation' might play a role as well, predicting a difference in amplitude opposite to the recorded one. As 'certainty' is multiplied with the other operands, while 'resource allocation' is added, it has a larger influence on the amplitude than 'resource allocation'.

This sub-experiment served as a control measure to test the setup of the experiment. Since we were able to find all expected components, we can conclude that the parameters for our recording and analysis are suitable for the planned analyses in the experimental sub-experiments. Moreover, the results of this sub-experiment indicate that the participants did not suffer from any auditory problems, understood the task and responded as expected.

4.2. Influence of degree of deviance in phoneme processing

The first speech-related sub-experiment was the 'auditory syllable' sub-experiment. Previous ERP research on phoneme processing concentrated on automatic, unconscious processing, using passive oddball designs (Sams et al., 1990; Kraus et al., 1992; Sharma et al., 1993; Sharma & Dorman, 1999) and does therefore not add to the understanding of attention-related processes. Lawson & Gaillard (1981) conducted an active oddball study, however they used rather large contrasts between standard and deviant. In the current study, we chose stimuli that differed in only one dimension ('place of articulation'), but to different degrees. The deviants, /ka/ and /ta/, were presented together with the standard, /pa/. Based on the TRACE model (McClelland & Elman, 1986), it was postulated that the difference between /pa/ and /ta/ is smaller than the difference between /pa/ and /ka/, when considering the overall difference of the three features 'acute', 'burst' and 'diffuse'. The phonemes /p/, /t/, and /k/ differ, for example, in their burst qualities: /p/ has a fairly faint burst that is scattered over a wide frequency range. The burst of /t/ is in a rather high frequency range and somewhat more intense than the one of /p/. The burst related to /k/ is in a middle frequency range and most intense. It is also longer than that of /t/, which in turn is longer than that of /p/ (Ladefoged, 2001). Therefore, /p/ and /t/ are closer to each other regarding the burst qualities than /k/ and /p/. This is not reflected in the behavioral results: there was no difference in accuracy nor in response times between the two deviants.

No MMN was found in this sub-experiment. In the designated time window, there was no difference between the deviants and the standard. How-

ever, there was a significant difference between 200 and 240ms (the N2b time window). Since this negativity started somewhat earlier than the time window we analyzed, it is most likely that it is a non-differentiated N2, consisting of both MMN and N2b influences. Sams et al. (1990) also found an N2 for phonemic contrasts in an active oddball design with no distinction into frontal early MMN and central later N2b. We found no difference in the responses to the two deviants in the first two time windows. Thus, in the phase of automatic processing, both contrasts are processed equivalently. When comparing differences in one and two dimensions, Lawson & Gaillard (1981), however, found an influence on the N2 amplitude: the amplitude was higher for a difference in two phonemic dimension than for a difference in one. In the current study the two contrasts were within the same phonetic dimension, place of articulation.

A large P3 caused by attention-related processes involved in difference detection was found. The amplitude was larger for the deviant /ta/ than for the deviant /ka/. As explained above, the distance between /pa/ and /ta/ is smaller than that between /pa/ and /ka/. This means that the smaller difference elicited the larger amplitude, unlike in the 'tones' sub-experiment.

As explained above, the amplitude of the P3 depends on three factors: probability, equivocation and resource allocation (Johnson, 1984, 1986). As for the 'tones' sub-experiment, the probability of both deviants was equal. However, in this sub-experiment also the equivocation did not differ, as the response times were almost identical. Therefore, only the third factor, 'resource allocation', distinguishes the two deviants. This means that the contrast needing most resources elicits the highest amplitude and that is in this

case the deviant /ta/. 'Resource allocation' is a rather unspecific term that does not explain which neuropsychological process actually causes the influence on the amplitude. In this case, the 'resource' in question can probably best be described as the attention necessary to detect different types of deviants.

In the visual sub-experiment, the deviants elicited more positive responses than the standard in all three time windows with no sign of a mismatch negativity as in the auditory modality. The onset of this positivity is rather early, starting in what is considered to be the pre-attentive time window for auditory processing.

The ERP response in the P3 time window showed a positivity which was larger for /ta/ than for /ka/. Behaviorally, there is a difference between both deviants as well: the accuracy was higher and the reaction time shorter for the deviant /ta/ than for the deviant /ka/, indicating a higher certainty. This is comparable to what was found for the tones: the stimulus with the higher uncertainty evokes the larger component. It differs, however, from the results for auditory syllables, where no behavioral difference between the deviants was found.

4.3. Audiovisual processing

When studying audiovisual processing, not only brain responses to the auditory and the visual information are recorded but also activity related to the integration of both. These effects were studied in the audiovisual sub-experiment. In order to investigate the process of integration, we looked at McGurk stimuli and compared them to audiovisual congruent stimuli. Furthermore, we compared the correlates of audiovisual perception to those of

auditory perception to determine whether the beneficial effect of audiovisual speech is represented in neural activity as well.

The behavioral results of the audiovisual sub-experiment show that /ta/ was detected as a deviant more easily than /ka/. The accuracy was lower and the response time was longer for the latter. This resembles the findings in the visual sub-experiment rather than the auditory results, in which both were equally detectable, and probably reflects the contribution of visual information.

The onset of the visual difference and the auditory difference were 200ms apart. Therefore, visual cues were picked-up earlier and the components evoked by the visual mismatch can overlay those related to the auditory mismatch detection. In an active design, due to the invested attention, a positivity related to the visual difference is expected. This is what we found and reported above for the visual syllables. Because the onset of the auditory difference is 200ms later than the onset of the visual difference, any auditory negativity might be covered by the large visual positivity. Therefore, we subtracted the visual activity from the audiovisual and then compared the remaining auditory response to the auditory syllables. Therefore, in the discussions below 'audiovisual' refers to the audiovisual activity after subtraction of the visual activity.

4.3.1. Brain correlates of the McGurk effect

The McGurk effect is a special case of audiovisual integration, as even though the auditory and the visual input do not match integration takes place. Earlier studies using the ERP paradigm concentrated on showing that the McGurk effect can elicit an MMN (Colin et al., 2002, 2004; Saint-

Amour et al., 2007). As the MMN is regarded to react only to auditory mismatch detection, the authors assumed that the auditory perception is altered by the misleading visual information. However, these studies were all carried out with a passive oddball design, providing no information about the actual perception of the participants and limiting the analysis to components related to automatic processing.

In the current study we extended the paradigm and compared the components elicited by McGurk stimuli to those elicited by congruent audiovisual stimuli. Only the McGurk stimuli, which participants perceived as deviant (indicated by pushing a button), were included in the analysis. It is therefore clear that the participants did not perceive the auditory part of the McGurk syllable. They could either have an altered auditory perception or have reacted to the visual difference. As explained above, we subtracted the activity recorded for the visual syllables from the activity of the audiovisual syllables. For the McGurk stimuli, this means that the visual activity of the deviant /ka/ was subtracted from the measured audiovisual activity. Therefore, only the auditory part, which did not differ from the standard, was taken into the analysis.

The McGurk stimuli elicited a more negative wave than the standard stimuli in all analyzed time windows. In the latest time window, this activity was more negative than that related to congruent audiovisual deviants. The difference with the standards is noteworthy because standards and McGurk items did not differ auditorily. The only differences were in the visual part and in the result of audiovisual integration. As the visual activity had been subtracted, the standard and the McGurk items were actually physically the

same. The response differed, however, substantially. As audiovisual processing was necessary for all stimulus types in this sub-experiment, the additional activation must be specifically caused by the integration of non-matching information. Möttönen et al. (2002) also found a difference between congruent and incongruent deviants in their (passive oddball) MEG study, which was limited to the right hemisphere. Processing of incongruent stimuli might be more complicated than processing of congruent stimuli: for congruent stimuli, both auditory and visual information contribute to the identification of the correct phoneme. For the incongruent stimuli, however, contradictory information is received, demanding more effort to select the matching phoneme and increase uncertainty about the given response.

4.3.2. Effects of audiovisual integration

In order to measure the effect of integration on processing, difference waves between the deviants and standards were calculated for the corrected audiovisual and for the auditory activity. As the audiovisual brain responses were corrected for the visual activity, any difference between the two presentation modalities should be due to the integration effects. Another possible explanation is that other effects, such as saliency or ease of difference detection play a role in the difference between audiovisual processing and auditory processing.

As shown in Figure 7, the observed negativity has an earlier onset for the 'corrected audiovisual' than the 'auditory' stimuli. Möttönen et al. (2002) found comparable results in an MEG study, in which they reported a shortened latency for audiovisual differences (compared to visual differences). This is another indication that audiovisual information facilitates processing.

There is a faster response to the mismatch detection, when both auditory and visual information are present than when only the auditory information is provided. The direct influence of the visual information on the activity had been subtracted from the wave. Therefore, the remaining effects were due to audiovisual integration processes or reflect differences in saliency or ease of deviant detection. Regardless of the chosen explanation, it can be concluded that the activity for the audiovisual stimuli is more than a mere addition of auditory and visual activation.

The positivity in the P3 time window was much larger for the purely auditory stimuli than for the 'corrected audiovisual' stimuli. This could reflect the difference in required 'resources' (such as attention): for the processing of audiovisual differences less attention is required. Therefore the amplitude of the P3 is smaller. This implies that the integration of audiovisual information does not come with a 'cost', but rather eases processing, resulting in a smaller P3 amplitude. Another possible explanation is based on the fact that the visual information leads the auditory by 200ms. Hence, the mismatch is first detected visually eliciting a P3. Once this has happened, there is no necessity to do any further mismatch processing based on the auditory input. Therefore, no P3 related to the auditory part of the input is recorded. Furthermore, the smaller P3 amplitude could be due to an overlay of a negativity related to integration. This would, however, imply that integration is a rather late process. This does not seem to be the case, as the effects seem to occur earlier in the audiovisual modality.

5. Conclusions

In this paper we addressed three main issues. We investigated whether the components representing automatic and conscious processing differed between two distinct contrasts (/pa/ vs. /ta/ and /pa/ vs. /ka/) as they do for tones, which we used as a control measure here. This was not the case for the components related to automatic processing, but it was true for the P3, representing conscious mismatch detection. We concluded that the smaller the difference is, the more attention is needed to detect the deviant and the larger the amplitude becomes. Since this was only tested with one contrasts, some caution needs to be paid in drawing conclusions. It is recommendable to investigate more contrasts to confirm these findings.

The second issue we addressed was whether the processing of McGurk stimuli differed from the processing of congruent audiovisual material. The McGurk stimuli elicited a more negative waveform than congruent standards and deviants, which cannot be explained by physical differences. More difficult integration is, therefore, the most likely explanation for this additional activity.

Finally, we addressed the effect of audiovisual integration. We investigated whether audiovisual processing provoked responses differing from those of a summation of auditory and visual processing. A comparison of activity due to auditory stimuli and activity due to audiovisual activity (after subtracting the visual activity) showed activation patterns differed, with a diminished P3 and a shorter latency for the audiovisual stimuli. This indicates that audiovisual integration facilitates processing.

The current study emphasized the influence of audiovisual processing on

comprehension. While it was known from behavioral studies that additional visual information facilitates (Sumby & Pollack, 1954; Reisberg et al., 1987) and influences (McGurk & MacDonald, 1976) comprehension, the present results from electrophysiological measures strengthen the claim that for audiovisual processing the whole is more than the sum of its parts.

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*, 197–207.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer [computer progam]. Version 5.2.17, retrieved 6 February 2009 from http://www.praat.org/.

Campbell, R. (1988). Tracing lip movements: Making speech visible. *Visible Language*, *22*, 32–57.

Campbell, R. (1990). Lipreading, neuropsychology, and immediate memory. In G. Vallar, & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 268–286). Cambridge: Cambridge University Press.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology*, *115*, 1989–2000.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A

phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*, 131–143.

Deacon, D., Nousak, J. M., Pilotti, M., Ritter, W., & Yang, C.-M. (1998). Automatic change detection: Does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology*, *35*, 413–419.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*, 259–273.

Johnson, R. (1984). P300: A model of the variables controlling its amplitude. *Annals of the New York Academy of Sciences*, *425*, 223–229.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*, 367–384.

Kraus, N., McGee, T., Sharma, A. M. A., Carrell, T., & Nicol, T. B. S. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear and Hearing*, *13*, 158–164. Using Smart Source Parsing.

Ladefoged, P. (2001). *Vowels and Consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishers Ltd.

Lawson, E. A., & Gaillard, A. W. K. (1981). Mismatch negativity in a phonetic discrimination task. *Biological Psychology*, *13*, 281–288.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, *16*, 68–89. Article Using Smart Source Parsing.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGee, T. J., King, C., Tremblay, K., Nicol, T. G., Cunningham, J., & Kraus, N. (2001). Long-term habituation of the speech-elicited mismatch negativity. *Psychophysiology*, *38*, 653–658.

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

Morin, R. E., & Forrin, B. (1963). Response equivocation and reaction time. *Journal of Experimental Psychology*, *66*, 30–36.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, *13*, 417–425.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.

Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology*, *14*, 53–98.

Novak, G., Ritter, W., & Vaughan Jr., H. G. (1992). Mismatch detection and the latency of temporal judgments. *Psychophysiology*, *29*, 398–411.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). Mmn in the visual modality: a review. *Biological Psychology*, *63*, 199–236.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli. In B. Dodd, & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 97–114). London: Lawrence Erlbaum.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*, 587–597.

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience*, *2*, 344–357.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*, 437–448.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*, 1078–1083.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*, 64–71.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212–215.

Figures

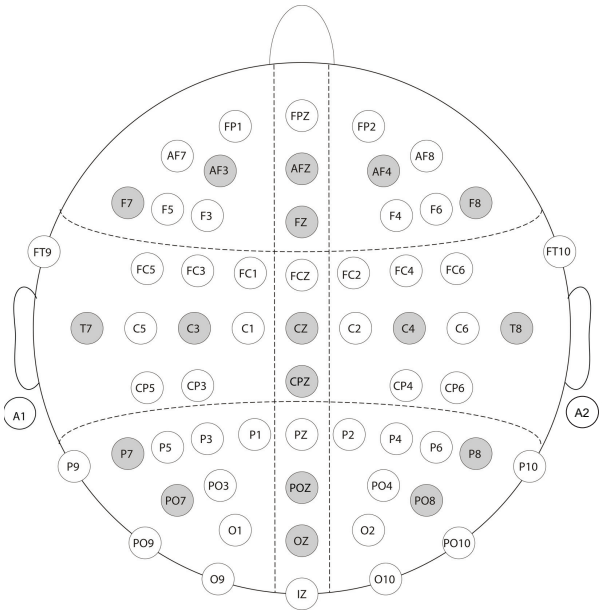

Figure 1: Position of the electrodes used in the current study. Electrodes marked in gray were used in the analyses. Regions of interest are indicated by dotted lines for the factors frontality (frontal, central, occipital) and laterality (left, midline, right).

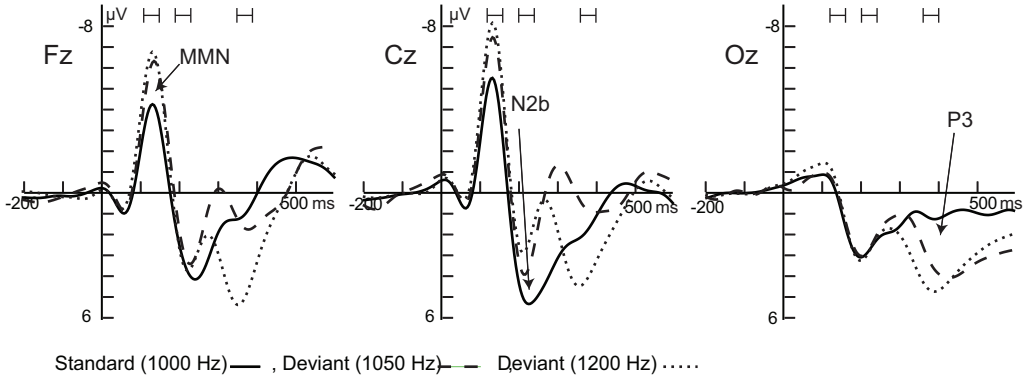

Figure 2: ERP of the three stimuli in the tones condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the Figure and labeled 'I1', 'I2' and 'I3'.

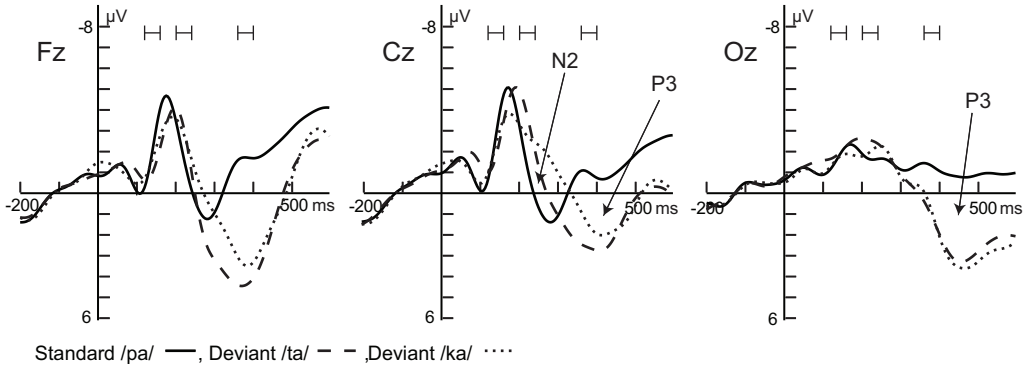

Figure 3: ERP of the three stimuli in the auditory syllables condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the Figure and labeled 'I1', 'I2' and 'I3'.

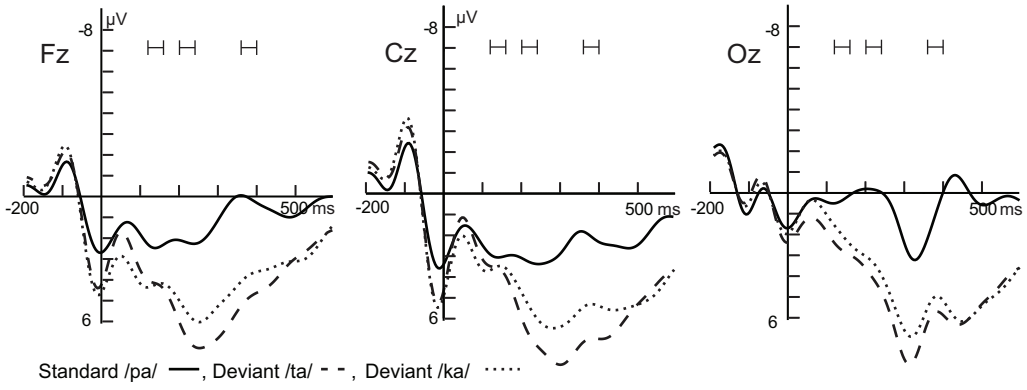

Figure 4: ERP of the three stimuli in the visual syllables condition. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the Figure and labeled 'I1', 'I2' and 'I3'.

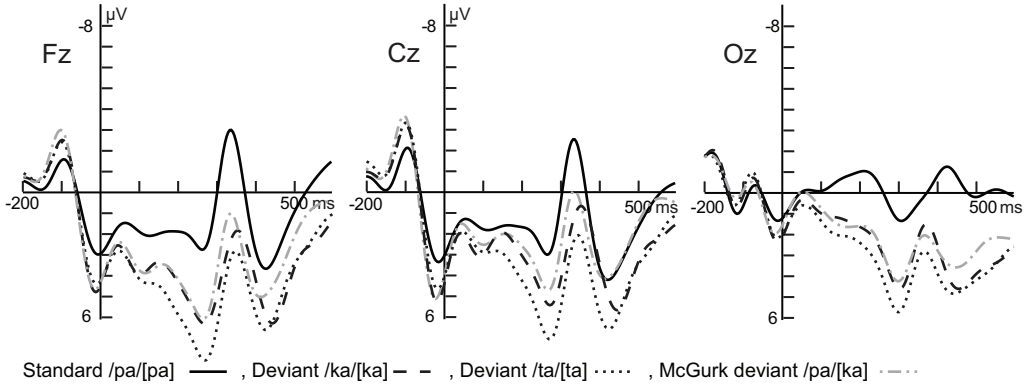

Figure 5: Uncorrected waves for all four stimuli in the audiovisual syllables condition. Displayed is the activity at the three electrodes Fz, Cz and Oz.

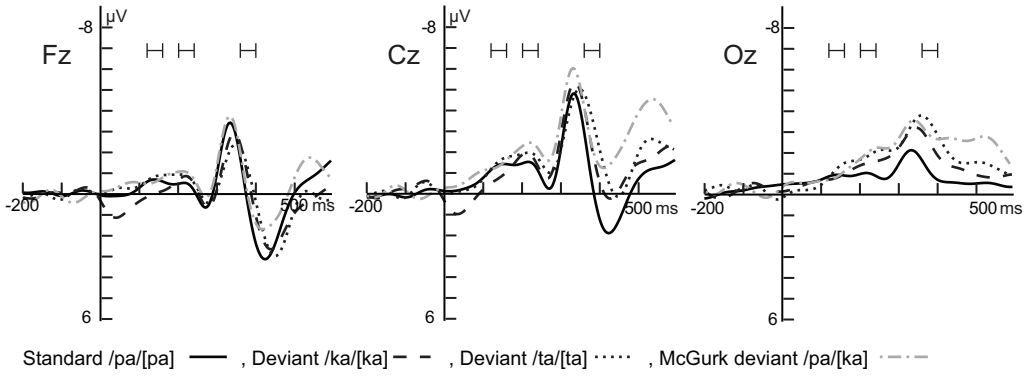

Figure 6: Difference waves (audiovisual - visual) for all four stimuli. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the Figure and labeled 'I1', 'I2' and 'I3'.

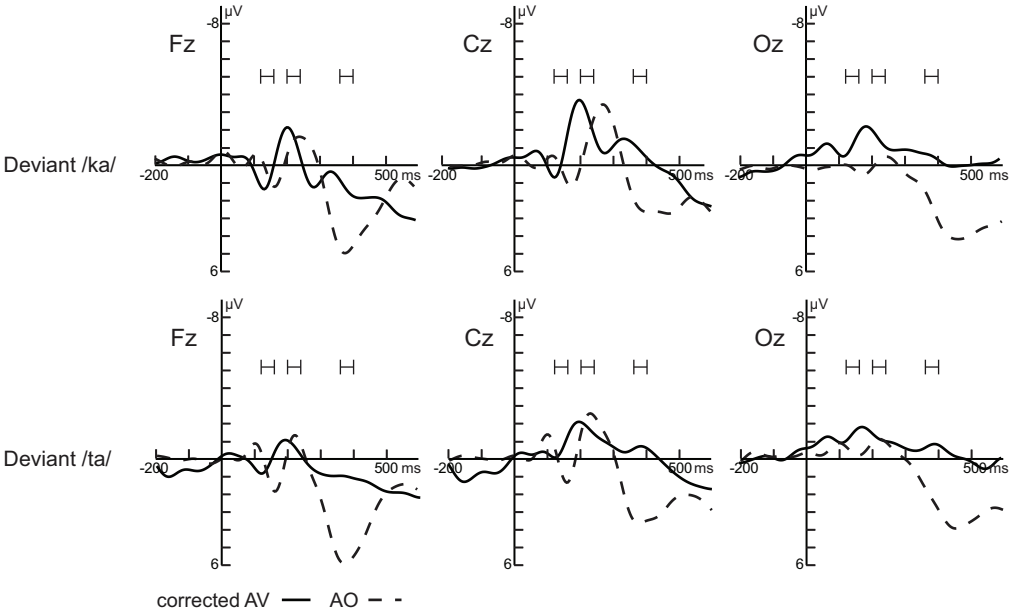

Figure 7: Difference waves (deviant-standard) for both deviants in the conditions with auditory presentation and the subtraction of visual presentation from audiovisual presentation. Displayed is the activity at the three electrodes Fz, Cz and Oz. The intervals chosen for statistical analyses are marked in the Figure and labeled 'I1', 'I2' and 'I3'.

Acknowledgements

This project was funded by an Ubbo Emmius Scholarship of the University of Groningen to the first author. Research of the second, third and fourth author was funded by the University of Groningen.

The authors gratefully acknowledge Callista Jippes for help with the preparation of the material and the assistants who helped with the recordings of the ERPs. Furthermore, they would like to thank all participants for volunteering to take part in this study. We also thank Lee Osterhout and an anonymous reviewer for their valuable comments on an earlier version of this paper.