

Title: Mismatch negativity (MMN) to speech sounds is modulated systematically by manual grip execution_

Authors: Mikko Tiainen^{*a}, Kaisa Tiippana^a, Petri Paavilainen^{a,b}, Martti Vainio^c, Lari Vainio^a

Affiliations:

^a Department of Psychology and Logopedics, University of Helsinki, Finland

^b Cognitive Brain Research Unit, University of Helsinki, Finland

^c Department of Modern Languages, University of Helsinki, Finland

- Corresponding Author:

Name: Mikko Tiainen

Address: Siltavuorenpenger 1A, room 427, 00014 University of Helsinki, Finland

E-mail: mikko.o.tiainen@helsinki.fi

Abstract

Manual actions and speech are connected: for example, grip execution can influence simultaneous vocalizations and vice versa. Our previous studies show that the consonant [k] is associated with the power grip and the consonant [t] with the precision grip. Here we studied whether the interaction between speech sounds and grips could operate already at a pre-attentive stage of auditory processing, reflected by the mismatch-negativity (MMN) component of the event-related potential (ERP). Participants executed power and precision grips according to visual cues while listening to syllable sequences consisting of [ke] and [te] utterances. The grips modulated the MMN amplitudes to these syllables in a systematic manner so that when the deviant was [ke], the MMN response was larger with a precision grip than with a power grip. There was a converse trend when the deviant was [te]. These results suggest that manual gestures and speech can interact already at a pre-attentive processing level of auditory perception, and show, for the first time that manual actions can systematically modulate the MMN.

Keywords: MMN; speech; action; action-perception; gestures

Introduction

Hand and mouth movements are interconnected. For example, while pronouncing a syllable, the mouth is opened wider when watching a large object being grasped than when watching a small object being grasped [1]. In addition, the so-called mouth-hand mimicry theories suggest that people tend to mimic concurrent hand movements with their mouth [2,3]. Indeed, young children [4], and also chimpanzees [5], have a tendency to perform mouth movements in an imitative manner with fine-motor hand movements.

We have previously shown that speech sounds and hand grips have specific connections (e.g. [6,7]). For example, the consonant [k] is associated with the power grip and the consonant [t] with the precision grip, so that grip execution is faster when an associated consonant is simultaneously pronounced or prepared. This congruency effect can also be observed in vocal reaction times [7]. Our findings provide evidence of the systematic nature of the connections between hand and mouth movements, and are thus in line with the mouth-hand mimicry theories as well as with the theories proposing that manual gestures might have contributed to the development of articulatory gestures (e.g. [8,9]). We have hypothesized that the articulations involving the tongue body, such as [k] could be considered an analogue of a power grip, where an object is grasped by pressing it against the palmar surface of the hand (e.g. [10]). In contrast, [t], for which the tip of the tongue is used, would be the analogue of a precision grip, where the tips of the index finger and thumb are utilized to pick up small objects [10].

We have also shown that the above-discussed articulation-grip connection works in a syllable categorization task [11]. In this study, participants were asked to prepare a power or a precision grip and then observe a noise-masked syllable. They executed the grip at the end of the syllable presentation and then reported whether the presented syllable was [ke] or [te]. When participants prepared/executed a power grip, they were more likely to categorize the presented syllable as [ke], and when a precision grip was prepared/executed, syllables were more often categorized as [te]. We found evidence that this categorization effect operates, at least in part, at the decision-making stage, but possible influences at earlier stages of perceptual processing could not be addressed thoroughly. In fact, it has been shown that stimulating mouth motor areas with transcranial magnetic stimulation (TMS) modulates speech perception systematically [12-14]. Therefore, if mouth and hand actions share overlapping processing networks, it is likely that the modulation of speech categorization by grips could also operate at early perceptual processing stages, in addition to the decision-making

stage. This idea is further supported by a recent functional magnetic resonance imaging (fMRI) study, which showed that grip execution modulates the activity in the auditory cortex while listening to meaningless non-speech sounds [15].

In the current event-related potential (ERP) study, we explored whether precision and power grip performance could have a systematic influence on perceptual processes related to heard syllables by utilizing the well-known ERP component, the mismatch negativity (MMN, e.g. [16]). The MMN is usually elicited in an oddball paradigm where occasional deviant stimuli (e.g. a tone of a different frequency) are presented in the midst of a stimulus sequence consisting of repetitive, physically identical “standard” stimuli. The MMN is seen in the deviant minus standard difference curve as a negativity peaking at around 100–200 ms after the deviant-stimulus onset. The MMN originates from the auditory cortex and is related to early stimulus processing (e.g. [17]). It has been theorized that representations of the regularities of auditory stimuli are stored in the auditory cortex and compared to incoming auditory information. The MMN represents the mismatch between the predicted regularity and the actual input ([18]). As MMN is observed even though no attention is paid to the auditory information, it is considered to reflect early and relatively automatic stages of auditory processing [16]. The MMN is also observed when using speech stimuli [16], and even with the McGurk illusion [19-21]. In the McGurk illusion, presenting incongruent visual and auditory speech stimuli alters the auditory perception [22]. These studies show that visual information can influence auditory processing at the pre-attentive level [19-21].

As the MMN to speech stimuli can be modulated by stimuli of another modality (i.e. visual speech), the question arises whether it could be influenced by other non-auditory means as well. As our previous study showed that grip actions can influence speech categorization [11], we hypothesized that grip execution could also specifically modulate MMN to speech stimuli when the grips are

associated with the speech stimuli. If this modulation could be observed, it would show for the first time that action-related processes modulate MMN patterns, and it would also strongly support the notion that manual actions interact with speech processing at an early stage and thus can influence the perceptual outcome. We used syllables [te] and [ke] as stimuli since they have been shown to be robustly associated with the precision and power grip, respectively [6,11]. We employed a classic oddball paradigm, where one syllable (e.g. [te]) was the standard stimulus, and the other one was the deviant (e.g. [ke]). In another block, the setting was reversed. The participants' task was to respond with appropriate grips to onscreen cues, while ignoring the sounds. Since grips were executed both during standard and deviant stimuli, we could create difference curves from trials that had exactly the same stimulus, but the grip context was different (i.e. physically the same speech stimulus as a standard vs. deviant, presented during two different types of grips). Thus, the possible effects on the MMNs could not be simply explained either by the two syllables producing different MMNs or by the brain responses being differentially contaminated by motor activity related to grip execution.

Material and methods

Participants

Twenty-one native Finnish-speaking volunteers participated in the study (19 women, 2 men, aged 23.6 ± 5.8 years). All were right-handed, and reported normal vision, hand motor functioning and hearing. This research was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences, and the participants gave their written informed consent.

Equipment, stimuli and procedure

The participant sat in a dimly-lit sound-attenuated room 100 cm in front of a 17" LCD-screen, holding two grip devices marked with green and blue tape in his/her right hand. The precision grip device was held between the index finger and thumb, and the power grip device with the rest of the fingers, squeezed against the palm of the hand (for more details on the grip devices see, e.g., Vainio et al., 2013).

The auditory stimuli were [ke] and [te] syllables, recorded from the speech of a Finnish woman and delivered binaurally through headphones. Both syllables were 160 ms long and edited to have an equal average f_0 of 241 Hz. The syllables were presented with an intensity of 50 dBA. The visual stimulus was a green or a blue circle (diameter 1.53 degrees of visual angle) shown in the centre of the screen.

The experiment used an oddball paradigm where one of the syllables was the standard stimulus ($P=.85$) and the other one was the deviant stimulus ($P=.15$). The participant's task was to respond to the circles when they appeared on the screen by squeezing the corresponding, colour-matching grip device. The participant was told to ignore the auditory stimuli. The syllables were presented with 500-ms interstimulus intervals. When a grip response was required, the visual go-signal for the response (green or blue circle) appeared on the screen 100 ms before the syllable was presented, so that the actual grip processing could start around the same time as the syllable processing, and the circle remained onscreen for 400 ms. The sequence of the deviant presentation was randomized except that two deviants were never presented in a row, and all grip responses were preceded by a standard stimulus. The experiment was split into two blocks where in one the standard was [ke] and the deviant was [te] and vice versa in the other one. The order of the blocks was balanced across

participants, as was the colour matching for the grip responses. Each block had six trial types: standard with no grip (2150 trials), standard with a power grip (300 trials), standard with a precision grip (300 trials), deviant with no grip (150 trials), deviant with a power grip (150 trials) and deviant with a precision grip (150 trials). There were more grip responses with standard stimuli to keep the task of the participant more involving and also to prevent associating grips directly with the deviant stimuli.

In total, one block had 3000 trials. Each block was split into three runs, and each run had three pauses whose length was participant-controlled. Between blocks there was a longer pause in order for the participant to rest properly. One run lasted for about 15 minutes, depending on the length of the pauses, and the entire experiment with preparations about 2.5 hours.

ERP recordings and statistical analysis

EEG was recorded with a 64-electrode cap (BioSemi ActiveTwo mk2, BioSemi B.V, The Netherlands), with three additional electrodes placed at the right and left mastoids and at the tip of the nose. Vertical electro-oculogram (EOG) was recorded with an electrode placed below the left eye and horizontal EOG with an electrode placed at the outer canthus of the left eye. The sampling rate was 512 Hz resulting in a dynamic range of DC–204Hz (2/5 of the sampling rate).

All ERP analyses were done using the nose electrode as the reference. The EEG data were first band-pass filtered (0.5–40 Hz). Independent component analysis (ICA) was used to identify and remove the components that most contributed to the eye blink signal. Before ICA, channels with bad signal were removed for each participant. The runica-algorithm was used for the ICA.

Removable components were chosen by manually comparing the components to the EOG signal. At

most two components were removed for each participant. After ICA, the EEG epochs related to individual trials were extracted, starting 100 ms before and lasting 500 ms after the onset of the auditory stimulus. Only correct-grip trials were used for the ERP averaging, which resulted in the rejection of 2.4% of the trials. Trials that had voltage changes exceeding $\pm 100 \mu\text{V}$ were also removed from the averaging. Including the rejected grip trials, in total 15.7% of the trials were rejected. One participant's data had to be completely rejected due to technical difficulties during recording. The remaining participants had at least 73 trials for each trial type (on average, 128 trials on deviant-grip trials). The 100-ms prestimulus time preceding the onset of the auditory stimulus was used as the baseline for amplitude measurements.

Difference curves were calculated by subtracting the ERPs to the standard trials from the corresponding deviant-trial ERPs (e.g. deviant [ke] with power grip minus standard [ke] with power grip). Consequently, the two trial types for each difference curve were physically identical (same grip and syllable), only the context was different (i.e. in one trial the syllable acted as standard stimulus and in the other it was the deviant). Time windows for the statistical analyses were selected according to visual observation, based on the most notable negativity peaks during the typical occurrence times of the MMN responses (100-200 ms) and the approximate duration of the negativities. According to these criteria, the time windows were 150–250 ms after auditory stimulus onset for [ke] and 100–200 ms for [te] (see Fig. 1 for example trials). The mean amplitudes during the selected time windows were calculated for the electrodes F3, F1, Fz, FC3, FC1, FCz, C3, C1 and Cz. Left-side electrodes were selected since in speech-related tasks, the MMN response originates mainly from the left hemisphere (e.g. [23]). Also, some participants had poor signals from the right-side electrodes. For the selected electrodes, a repeated-measures ANOVA ($2 \times 2 \times 3 \times 3$ design) was performed with the factors *deviant syllable* ([ke] and [te]), *grip* (power and precision), *electrode row* (i.e., the frontal-central axis: F, FC and C) and *electrode column* (i.e., the

left-midline axis: 3, 1 and z). We only compared the trials where grip responses were executed, since we were interested in the possible MMN modulations related to the two grips. Greenhouse-Geisser correction was used when the sphericity assumption was violated. Partial eta-squared served as the effect size estimate. Pairwise comparisons were Bonferroni-corrected.

Results

The difference curves (Fig. 1, bottom) show an MMN on each trial type, peaking during 150–250 ms time window after stimulus onset in the deviant [ke] block and during 100–200 ms time window in the deviant [te] block. The isopotential map for deviant [ke] precision grip trials (Fig. 1, top) shows the typical MMN distribution: the largest amplitudes were recorded at the fronto-central areas over the left hemisphere and the MMN inverted polarity at the electrodes below the level of the Sylvian fissure (e.g. the mastoids), due to its generator sources at the auditory cortices. The MMNs to deviant [ke] were larger than those to deviant [te]. Interestingly, the grip modulated the MMN amplitude so that MMN to deviant [ke] was larger with precision grip than with power grip. In the case of deviant [te], the effect was reversed (Fig. 1).

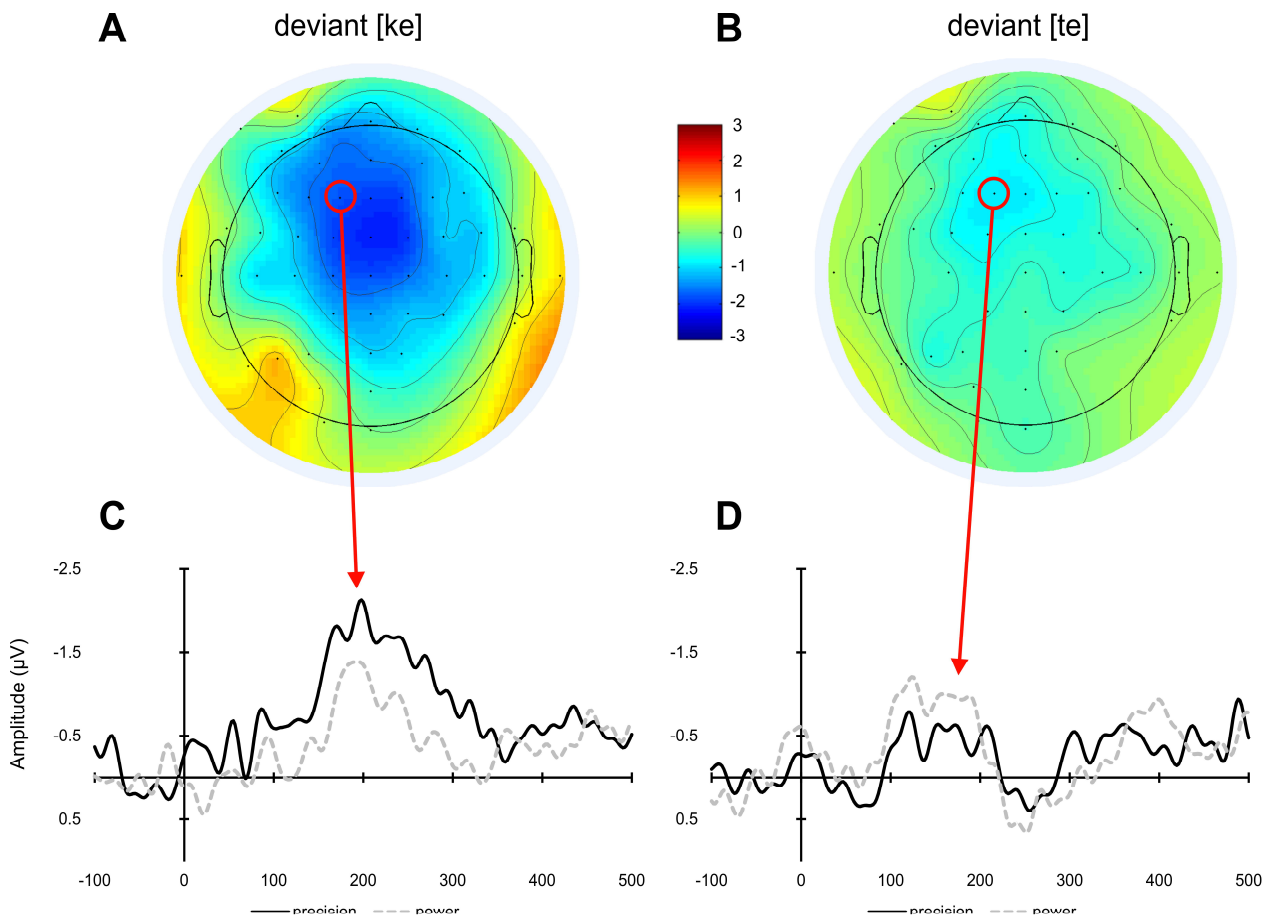


Figure 1. A: Isopotential map for deviant [ke] precision grip trials during a 180–220 ms time window, which coincided with the largest negativity peak. **B:** Isopotential map for deviant [te] power grip trials during a 140–180 ms time window. **C:** Example of ERP grand average difference curves on electrode F1 from the deviant [ke] block ($n = 20$). **D:** Example of ERP grand average difference curves on electrode F1 from the deviant [te] block ($n = 20$).

The ANOVA of the selected 3×3 electrode array revealed that the only significant effects were the main effect of column [$F(2,38) = 15.66, p < .001, \eta^2 = .45, \text{power} > .99$] and the interaction between syllable and column [$F(2,38) = 5.31, p = .024, \eta^2 = .22, \text{power} = .81$]. The MMN on the leftmost column 3 ($-0.738 \mu\text{V}$) was weaker than on the 1 ($-0.980 \mu\text{V}, p < .001$) and z ($-0.965 \mu\text{V}, p = .006$) columns. Pairwise analysis showed that the difference between columns was significant only for the

[ke] syllable. The interaction between syllable and grip approached significance [$F(1,19) = 3.99$, $p = .060$, $\eta^2 = .17$, power = .47].

Since the MMN peaks were the clearest on the frontal F-row electrodes, we performed an additional ANOVA for only the F row with column factor changed to electrode (F3, F1 and Fz; see Table 1 for the MMN amplitudes). There was a significant main effect of electrode [$F(2,38) = 7.18$, $p = .002$, $\eta^2 = .27$, power = .91], the MMN amplitude was weaker on the leftmost electrode F3 than on F1 or Fz. Importantly, the syllable-grip interaction was also significant [$F(1,19) = 6.29$, $p = .021$, $\eta^2 = .25$, power = .66]. The MMN was larger with [ke] for precision grip than for power grip (-1.662 μV vs. -0.864 μV , $p = .042$) and there was a converse, although nonsignificant, trend with [te] (power grip -0.439 μV vs. precision grip -0.882 μV , $p = .132$). The difference between [ke] and [te] was significant for precision grip ($p = .037$), the MMN was larger for [ke] than [te]. Difference between syllables was not significant for power grip ($p = .972$).

Table 1.

Average MMN amplitudes in μV (standard deviations in parentheses) for each trial type for the frontal electrodes used in the analyses.

Electrode	deviant [ke] (150-250 ms)		deviant [te] (100-200 ms)	
	precision	power	precision	power
F3	-1.52 (1.37)	-0.69 (1.44)	-0.39 (1.54)	-0.80 (1.55)
F1	-1.68 (1.58)	-0.98 (1.39)	-0.49 (1.50)	-0.94 (1.70)
Fz	-1.79 (1.67)	-0.93 (1.72)	-0.44 (1.44)	-0.91 (1.61)

We continued to investigate the deviant [te] conditions for possible explanations why the grip-syllable interaction was not significant with the deviant [te]. A one-sample t-test on the F1 electrode – where largest MMNs were observed – revealed that the deviant [te] conditions did not significantly differ from zero with precision grip ($t(19) = -1.449$, $p = .164$) or no grip ($t(19) = 0.030$, $p = .976$), but did with power grip ($t(19) = -2.469$, $p = .023$). In contrast, the deviant [ke] conditions

all differed from zero on electrode F1 (all p 's < .05). Obviously, for some reason, the physical parameters of the [te] stimulus were not optimal for eliciting strong MMN responses.

Discussion

In the current study, we created a novel paradigm utilizing the MMN to explore the possible influence of executing different hand grips on speech processing. We found an interaction between syllable and grip type. When the deviant syllable was [ke] and a precision grip was executed, the MMN was larger than when a power grip was executed. This pattern reversed (although non-significantly) when the deviant syllable was [te], so that the MMN tended to be larger with a power grip than with a precision grip.

Previously, we found that syllables (presented either auditorily, audiovisually or visually) are more likely categorized as [ke] when preparing a power grip and as [te] when preparing a precision grip [11]. We were, however, unable to determine whether perceptual processes were also affected. The current MMN results suggest that manual actions can modulate speech processing at a pre-attentive level of perception. The motor theory of speech perception [24] suggests that the perception of speech is partially formed by simulating the heard speech in the listener's own speech-motor system. Our results are in line with this view, but expand it by suggesting that manual processes can also contribute to speech perception. The current results also extend the view so that motor activity during the perception phase affected the speech processing already at a pre-attentive level.

Due to the explorative nature of the study, we did not have specific hypotheses about how the MMN amplitude would be affected by the different grips. We speculate that since the precision grip is associated with [te] [6,11], the precision grip performance boosts the processing of standard [te],

which in turn amplifies the mismatch between the standard [te] and the deviant [ke]. In contrast, since the power grip is associated with [ke] [6,11], the power grip boosts the processing of the standard [ke], amplifying the mismatch between the standard [ke] and the deviant [te]. Even though the difference between grips was not significant for deviant [te], the fact that the MMN component on F1 electrode differed significantly from zero only when a power grip was executed is in line with the view that power grip execution made the [te] syllable more different from [ke]. That is, the power grip may have amplified the difference between the deviant [te] and the standard [ke] so that a significant MMN response emerged.

The reason why the comparison between grips with the deviant [te] was not significant could be related to the fact that the MMN in general was weaker for the [te] deviants. Some features of the stimuli made [ke] a more effective deviant than [te]. Consequently, it is possible that the grip difference would have been significant if the deviant [te] had produced a stronger MMN.

The asymmetries in the interaction could be also interpreted from another perspective. The difference between [ke] and [te] was significant for the precision grip but not for the power grip. It is possible that the power grip did not sufficiently influence the syllable processing that it could have significantly modulated MMN patterns between the syllables. This could be related to the fact that the precision and power grips were both performed with the same hand. However, the precision grip was executed in a natural manner by using the index finger and the thumb, whereas the power grip was performed somewhat unnaturally by flexing only the little, ring and middle finger. This question should be properly addressed in future studies, but at least the precision grip seems to consistently modulate auditory processes related to perception of the consonant congruent with the grip.

Conclusions

For the first time, hand motor actions were shown to influence the early pre-attentive processing of speech in the auditory cortex, as indicated by the MMN modulation. Thus, MMN can be modulated by information from different sources such as motor (the present study) and visual speech [19-21]. Our results also complement those of Wikman et al. [15], who showed that grip execution while listening to non-speech stimuli modulates auditory cortex activity. According to the current findings, grip execution can also modulate auditory cortex processes related to specific speech stimuli congruent with the grip.

Acknowledgements

Funding: This work was supported by the Academy of Finland [grant number 265610]; and the Alfred Kordelin Foundation [grant number 160417]. We would like to thank Ilkka Muukkonen for his help in collecting the data.

References

- [1] M. Gentilucci, Grasp observation influences speech production, *Eur. J. Neurosci.* 17 (2003) 179–184.
- [2] R. Paget, *Human speech*, Harcourt Brace, New York, 1930.
- [3] G.W. Hewes, Primate Communication and the Gestural Origin of Language, *Curr. Anthropol.* (1973) 5–24.
- [4] G.S. Forrester, A. Rodriguez, Slip of the tongue: Implications for evolution and language development, *Cognition* 141 (2015) 103–111.

- [5] G.S Waters, R.S. Fouts, Sympathetic mouth movements accompanying fine motor movements in chimpanzees (*Pan troglodytes*) with implications toward the evolution of language, *Neurol. Res.* 24 (2002) 174–180.
- [6] L. Vainio, M. Schulman, K. Tiippana, M. Vainio, Effect of syllable articulation on precision and power grip performance, *PloS one* 8 (2013) e53061.
- [7] M. Tiainen, K. Tiippana, M. Vainio, N. Komeilipoor, L. Vainio, Interaction in planning vocalisations and grasping, *Q. J. Exp. Psychol-a.* 70 (2016) 1590–1602.
- [8] M.A. Arbib, From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics, *Behav. Brain Sci.* 28 (2005) 105–124.
- [9] M. Gentilucci, M.C. Corballis, From manual gesture to speech: a gradual transition, *Neurosci. Biobehav. R.* 30 (2006) 949–960.
- [10] L. Vainio, M. Tiainen, K. Tiippana, M. Vainio, Shared processing of planning articulatory gestures and grasping, *Exp. Brain Res.* 232 (2014) 2359–2368.
- [11] M. Tiainen, K. Tiippana, M. Vainio, T. Peromaa, N. Komeilipoor, L. Vainio, Selective Influences of Precision and Power Grips on Speech Categorization, *PloS one* 11 (2016) e0151688.
- [12] A. D’Ausilio, F. Pulvermüller, P. Salmas, I. Bufalari, C. Begliomini, L. Fadiga, The motor somatotopy of speech perception, *Curr. Biol.* 19 (2009) 381–385.
- [13] R. Möttönen, K.E. Watkins, Motor representations of articulators contribute to categorical perception of speech sounds, *J. Neurosci.* 29 (2009) 9819–9825.
- [14] E.H. Smalle, J. Rogers, R. Möttönen, Dissociating contributions of the motor cortex to speech perception and response bias by using transcranial magnetic stimulation, *Cereb. Cortex* 25 (2015) 3690–3698.

- [15] P.A. Wikman, L. Vainio, T. Rinne, The effect of precision and power grips on activations in human auditory cortex, *Front. Neurosci.* 9 (2015).
- [16] R. Näätänen, P. Paavilainen, T. Rinne, K. Alho, The mismatch negativity (MMN) in basic research of central auditory processing: A review, *Clin. Neurophysiol.* 118 (2007) 2544–2590.
- [17] K. Alho, Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes, *Ear Hearing* 16 (1995) 38–51.
- [18] I. Winkler, Interpreting the mismatch negativity, *J. Psychophysiol.* 21 (2007) 147–163.
- [19] M. Sams, R. Aulanko, M. Hämäläinen, R. Hari, O.V. Lounasmaa, S.T. Lu, J. Simola, Seeing speech: visual information from lip movements modifies activity in the human auditory cortex, *Neurosci. Lett.* 127 (1991) 141–145.
- [20] C. Colin, M. Radeau, A. Soquet, D. Demolin, F. Colin, P. Deltenre, Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory, *Clin. Neurophysiol.* 113 (2002) 495–506.
- [21] R. Möttönen, C.M. Krause, K. Tiippana, M. Sams, Processing of changes in visual speech in the human auditory cortex, *Cognitive Brain Res.* 13 (2002) 417–425.
- [22] H. McGurk, J. MacDonald, Hearing lips and seeing voices, *Nature* 264 (1976) 746–748.
- [23] R. Näätänen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R.J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, K. Alho, Language-specific phoneme representations revealed by electric and magnetic brain responses, *Nature* 385 (1997) 432–434.
- [24] A.M. Liberman, F.S. Cooper, D.P. Shankweiler, M. Studdert-Kennedy, Perception of the speech code, *Psychol. Rev.* 74 (1967) 431–461.