Title: Attention modulates somatosensory influences in passive speech listening

Authors: Judith Schmitz¹*, Begoña Díaz¹, and Nuria Sebastian-Galles¹

Affiliations: ¹ Center for Brain and Cognition, Department of Technology, University Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain

E-mail addresses: judith.schmitz@upf.edu, begona.diaz@upf.edu,
nuria.sebastian@upf.edu

*Correspondence: Judith Schmitz
Center for Brain and Cognition
Universitat Pompeu Fabra,
Roc Boronat 138, 08018 Barcelona, Spain
+34 93 542 2629
judith.schmitz@upf.edu

Running title: Somatosensory influence in speech listening

Number of words: 8,890

Acknowledgments: We thank Robert Frank de Menezes for help with the data acquisition and Mathilde Fort, Miguel Burgaleta and Pamela Miller for their helpful comments on the manuscript. This research was funded by grants from the European Commission Seventh Framework Programme (FP7/2007-2013): ERC grant agreement number 323961 (UNDER CONTROL); Cooperation grant agreement number 613465 – (AThEME); the Spanish Ministerio de Economía y Competitividad (PSI2012-34071) and the Catalan Government (SGR 2014–1210) awarded to NSG. She also received the “ICREA Acadèmia” prize for excellence in research, funded by the Generalitat de Catalunya. BD received funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007–2013) under REA grant agreement number 32867 and a postdoctoral fellowship from the Spanish Government (Juan de la Cierva fellowship JCI-2012-12678).
Abstract

Previous studies showed that manipulating the speech production system influenced speech perception. This influence was mediated by task difficulty, listening conditions, and attention. In the present study we investigated the specificity of a somatosensory manipulation—a spoon over the tongue—in passive listening. We measured the Mismatch Negativity (MMN) while participants listened to vowels that differ in their articulation—the tongue height—and familiarity—native and unknown vowels. The same participants heard the vowels in a spoon and no-spoon block. The order of the blocks was counterbalanced across participants. Results showed no effect of the spoon. Instead, starting with the spoon enhanced the MMN amplitude. A second experiment showed the same MMN enhancement for starting with a somatosensory manipulation applied to a non-articulator—the hand. This result suggests that starting a study with a somatosensory manipulation raises attention to the task.

Keywords

speech perception, speech production, speech processing, event-related potentials, mismatch negativity


**Introduction**

Hearing speech activates a large brain network that involves not only auditory areas but also production areas (e.g. motor and somatosensory), yet the contribution of these non auditory areas to speech processing is still under debate (Hickok, Houde, & Rong, 2011; Pulvermüller & Fadiga, 2010; Scott, McGettigan, & Eisner, 2009). While some theories assume that articulatory gestures are important for speech perception (Liberman & Mattingly, 1985; Möttönen & Watkins, 2012; Pulvermüller & Fadiga, 2010), other theories claim that speech perception is merely an auditory process, relying only on the analysis of the psychoacoustic signal (Diehl, Lotto, & Holt, 2004). Still others claim that the link between speech perception and production is not a necessary component of the speech perception system and is only activated when the task or listening conditions are difficult (Hickok et al., 2011; Schwartz, Basirat, Ménard, & Sato, 2012; Scott et al., 2009). The aim of the present research is to investigate the relation between the speech perception and production systems by studying if a somatosensory manipulation on a speech articulator—placing a spoon over the tongue—can influence speech discrimination of vowels differing in vowel height in passive listening, as measured by the brain electrophysiological response Mismatch Negativity (MMN), and whether the influence depends on the familiarity with the speech sounds. An additional goal is to better understand how somatosensory manipulations impact on the attentional system.

Studies supporting a critical role of the production system in speech processing showed that perception of phonemes is influenced by manipulating the articulatory muscle used during the production of the same phoneme, either by stimulating the area of a speech articulator (e.g. tongue or lips) in the primary motor cortex by means of transcranial magnetic stimulation (TMS) (D’Ausilio et al., 2009; Möttönen & Watkins, 2009) or by
directly constraining the gesture of the articulators (Ito, Tiede, & Ostry, 2009; Nasir & Ostry, 2009; Sato et al., 2011). For example, Ito et al. (2009) used a robotic device to stretch participants’ skin next to the lips upwards or downwards concurrent to the onset of the auditory presentation of words from the continuum between “head” and “had”. The two words are experienced by an upward (“head”) or a downward (“had”) stretching of the facial skin next to the lips respectively during articulation. The results showed that this somatosensory manipulation shifted the behavioral judgment of the participants in the same direction as the manipulation was applied. When the lip was stretched upwards, more words were identified as “head” and the downward movement shifted the identification to the word “had”. Importantly, the authors showed that this effect was specific to the motor constraint they applied: stretching the skin differently from the somatosensory experience of both vowels (i.e., backwards) or stretching the skin in a non-speech like manner (e.g. a twitch that was shorter than the sounds and faster than a speech movement) did not influence participants’ vowel categorization.

This articulatory specific influence of the speech production system on perception led to the interpretation that the link between the two systems is necessary for speech perception. Yet, studies using electroencephalography (EEG) or its magnetic counterpart, magnetoencephalography (MEG), revealed that the specificity of such interaction depended on attention (Möttönen, Dutton, & Watkins, 2013; Möttönen, van de Ven, & Watkins, 2014). Möttönen et al. (2013, 2014) assessed the MMN and auditory ERP components elicited by phonemes articulated with the lip or the tongue in attentive (active listening) and unattentive (passive listening) conditions after deactivating the lip representation in the motor cortex with rTMS. The MMN is a frontal negativity, with a reversed polarity in the mastoids, elicited at around 100-250
ms after a change between one frequent sound (standard) and an infrequent one (deviant) is detected (Duncan et al., 2009; Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). It is elicited by nonlinguistic (e.g., tones or music) and linguistic stimuli (e.g., phonemes or words). The amplitude and latency of the MMN are related to the amount of perceived change between the standard and the deviant (Amendo & Escera, 2000; Pakarinen et al., 2013) and is therefore considered an index of auditory discrimination sensitivity. The results first yielded that when attention was paid to the stimuli, the amplitude of the auditory component P50m in the left auditory cortex decreased for speech sounds articulated with the lips—thereby showing an articulator-specific influence on speech processing similar to previous TMS studies (e.g. D’Ausilio et al., 2009; Möttönen & Watkins, 2009). However, at a later time window, in both active and passive listening conditions, the amplitude of the auditory component P200m decreased bilaterally for both speech sounds articulated with the lips and with the tongue. The same articulatory unspecific decrease was found for the MMN (elicited in passive listening conditions). These results show a general influence of the production system on speech perception independently of the muscle used to articulate the perceived speech sounds. The authors explain these results by suggesting that attention regulates the specificity of motoric influences in speech perception. During active listening, the motor system modulates speech perception in an articulator-specific manner. During passive listening, the motor system might automatically model the state of all articulators to facilitate automatic speech perception.

Besides attention, another relevant factor modulating the effect of the production system in speech perception is task difficulty. This is in line with the claim that speech
perception does not necessarily involve the speech production system. According to Hickok et al. (2011), Scott et al. (2009), and Schwartz et al., (2012), the link between speech perception and production areas is only recruited during certain conditions, such as when the task is difficult or the listening conditions are not optimal. For example, D’Ausilio et al. (2009) and D’Ausilio, Bufalari, Salmas, and Fadiga (2012) used TMS to stimulate either the lip or tongue representation in the motor cortex while participants were listening to phonemes articulated with the lips and tongue when the phonemes were embedded in noise (difficult condition) or without any noise mask (easy condition). Results showed that TMS stimulation influenced the behavioral performance of the participants only in the difficult condition, when the sounds were embedded in noise. Similar results have also been found in fMRI studies (Osnes, Hugdahl, & Specht, 2011) and MEG studies (Alho et al., 2012). These studies showed that besides attention, task difficulty modulated the influence of the production areas during speech perception.

Previous studies have used different manipulations to increase the difficulty of the task, such as masking the speech (Alho et al., 2012; D’Ausilio et al., 2009; Osnes et al., 2011), using different speakers (Bartoli et al., 2015), using speech continua (Ito et al., 2009; Möttönen & Watkins, 2009; Nasir & Ostry, 2009), or using unknown speech sounds (Wilson & Iacoboni, 2006). A previous fMRI study has shown that the production system (i.e. premotor cortex) is activated when listening to non-native phonemes (Wilson & Iacoboni, 2006). Although the recruitment of the production system helps to perceive native speech in difficult tasks (e.g. higher accuracy and faster reaction times; Alho et al., 2012; D’Ausilio et al., 2009), for non-native speech the higher activation of the production system might not necessarily lead to better
performance. Wilson & Iacoboni, (2006) propose another explanation. According to these authors, the role of the motor system is to generate a model of the articulatory gestures of the perceived phoneme and to send this model to the auditory regions for comparison with the acoustic signal. For native phonemes this works well, as a fit for the perceived phoneme is found. However, for non-native phonemes, the motor system is not able to find a match for the unknown articulatory gesture. These unsuccessful tries would lead to higher activations of the motor system. Their study showed that familiarity with the articulatory gestures of phonemes decreased the processing cost in the production areas.

In summary, previous studies have shown that speech perception can be influenced by manipulating the production system (D’Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Nasir & Ostry, 2009; Sato et al., 2011), however this influence is dependent on attention and task difficulty. Attention modulates the specificity of the production influence in speech perception (Möttönen et al., 2013, 2014). Paying attention to the stimuli results in articulatory specific influences in speech perception while in passive listening conditions the position of all articulators is accessed, resulting in unspecific influences of the production system in speech perception. Task difficulty modulates if the production system gets engaged in speech perception with only difficult conditions, such as masked speech or listening to non-native phonemes, triggering the production system (Alho et al., 2012; D’Ausilio et al., 2012; Osnes et al., 2011; Wilson & Iacoboni, 2006). These studies show that it is still not resolved if the production system is a necessary component during speech perception. Furthermore, very little is known about the influence of the production system in passive listening conditions (Möttönen et al., 2013, 2014) and the perception of non-native phonemes
(Wilson & Iacoboni, 2006). The aim of this study is to investigate the specificity of the influence of the production system in speech perception in passive listening conditions for known and unknown phonemes. In line with the previous literature, we expect to find an unspecific influence of the production system on speech perception in passive listening conditions for native phonemes (Möttönen et al., 2013, 2014). For unknown phonemes we expect no influence of the production manipulation as proposed by Wilson and Iacoboni (2006), who claim that the production system cannot model the motor gestures for unknown phonemes.

**Experiment 1**

The goal of the first experiment was to test if a somatosensory manipulation—a spoon over the tongue—influences speech perception in passive listening conditions in an articulator specific manner and if this influence is modulated by the familiarity of the phonemes. The effect of a spoon was assessed on Spanish listeners MMN responses while participants passively listened to vowels that vary in tongue height when articulated. The Spanish native vowel /e/, articulated with a mid tongue height, was presented frequently, as the standard stimulus. The deviants were presented with a lower probability and were the native Spanish vowels /a/, /i/, /o/ and /u/, and the unknown Finnish vowels /ö/ and /y/. We expected the Finnish vowels to elicit a reliable MMN, since they are distant from the Spanish vowels, see Figure 1, and are easily discriminated from the Spanish vowels behaviorally, as tested with a discrimination task. The deviants /i/, /u/ and /y/ are articulated with a high tongue position and /a/, /ö/, /u/ are articulated with a low or mid tongue height. The somatosensory manipulation—the spoon over the tongue—blocks the tongue in a low position, similar to the articulation position for low vowels such as /a/, changing the natural position of the
tongue within the mouth to a lower position. Such a manipulation has been shown to fundamentally affect the speech production system leading to an impairment of speech discrimination in preverbal infants (Bruderer, Danielson, Kandhadai, & Werker, 2015). Bruderer et al. (2015) studied the discrimination abilities of 6-month-olds to a non-native, unknown (Hindi) speech contrast when the infants’ tongue was blocked by a flat teething toy as compared to having no teething toy or one that did not impede tongue movement. Only when the tongue was blocked, infants could not discriminate the non-native speech contrast. The authors concluded that sensorimotor information from speech articulators influence speech perception even in the absence of productive or perceptive experience with the speech sounds. Thus, we expect that the present somatosensory manipulation, i.e., the spoon over the tongue, will perturbate the speech production system.

The experiment consisted of two blocks in which the exact same speech stimuli and procedures were used with the exception of the somatosensory manipulation, i.e. with and without a spoon over the tongue, with the order of the blocks counterbalanced between participants. If during passive listening the production of all articulators is modeled, as claimed by Möttönen and colleagues (Möttönen et al., 2013, 2014), we expect that the spoon over the tongue affects the MMN amplitude for the native phonemes, independently of the vowel type (low vs. high tongue position). The spoon over the tongue could affect the MMN amplitude in two ways. The spoon over the tongue could activate the tongue muscle, leading to an increase in the MMN amplitude. This effect would be similar to a previous TMS study which showed that increasing the excitability of the tongue representation in the motor cortex with TMS enhanced the behavioral discrimination of speech sounds involving the tongue (D’Ausilio et al.,
Alternatively, the spoon over the tongue could inhibit the ability of the participants to simulate the motor response required to articulate the phonemes, thereby reducing the MMN amplitude. This effect would be similar to previous rTMS studies which showed that deactivating the lip representation in the motor cortex led to poorer behavioral performance and reduced MMN amplitudes of speech sounds articulated with the lips or speech sounds in general (Möttönen et al., 2013; Möttönen & Watkins, 2009). As for the unknown vowels, if the production system is not able to reconstruct motor gestures for unknown phonemes, as suggested by Wilson and Iacoboni (2006), the MMN elicited by the unknown phonemes should not be modulated by the spoon manipulation. To control that the order of the blocks (starting with the spoon or without the spoon) had no influence on the MMN we included this factor in the analysis.

**Methods**

**Participants**

Thirty-seven native Spanish listeners (23.84 years ± 3.82, 31 females) participated in the experiment. All participants were right handed, without hearing problems and none self reported being proficient in any languages whose phonological repertoire included the vowels /ö/ or /y/ (the unknown vowels used in the present study), in particular French and German. The participants were university students except for two participants who were vocational students. The study was conducted in accordance with the Declaration of Helsinki and the local ethical committee. Participants provided written informed consent and were paid for their participation (10 €/h). Data from one participant was excluded due to problems with the ocular correction and data from two participants was excluded in an outlier analysis (MMN amplitude more than 2.5
standard deviations from the mean in any of the MMN conditions) leaving a final sample of thirty-four participants.

**Stimuli**

Auditory stimuli were the five Spanish vowels /a/, /e/, /i/, /o/, /u/ and the two Finnish vowels /ö/ and /y/. The vowels were synthesized using the “Simplified Vowel Synthesis Interface” (Speech Research Lab, DuPont Hospital for Children and the University of Delaware, http://www.asel.udel.edu/speech/tutorials/synthesis/vowels.html). The synthesizer uses the Klatt synthesizer to create vowels. For all the vowels, the values for the first and second formant frequency (F1 and F2) as measured by Praat can be seen in Figure 1. For the Spanish vowels we used the values described in Martínez Celdrán and Fernández Planas (2007) based on the mean values of 5 male speakers of standard Spanish. For the Finnish vowel /ö/ we used the same values as have been used in Näätänen et al. (1997) and Díaz, Baus, Escera, Costa, & Sebastián-Gallès (2008). For the Finnish vowel /y/ we used the same values as in Cheour et al. (1998). The pitch (110 Hz) and the third to fifth formant frequencies were kept constant during synthesis for all vowels (F3: 2540 Hz, F4: 3300 Hz, F5: 3850 Hz). The duration of all phonemes was 200 ms including 10 ms of rise/fall times.

Because a reduced MMN is present when non-native vowels are perceived as similar to native (Näätänen et al., 1997; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005; Winkler et al., 1999), we ensured that native Spanish speakers can discriminate native (/a/, /e/, /i/, /o/, /u/) from unknown vowels (/ö/, /y/). An additional group of six native Spanish listeners (25.00 years ± 6.00, 3 females) with the same characteristics as described performed an AXB discrimination task. During the AXB task, two different
stimuli were presented in the first and third (last) position, while the second (middle) stimulus was either the same as the first one or the third one. Trials always included one unknown vowel, /ö/ or /y/, and one of the native vowels, /a/, /e/, /i/, /o/, or /u/ (e.g., /ö-ö-a/, /i-i-y/, etc...). In 80 trials, participants had to respond via button press whether the different vowel was the first or third element of the triplet. The participants were at the ceiling in discriminating the unknown vowels from the native vowels (accuracy of 99.375% ± 0.427).

Procedure

The experiment was carried out in an electrically shielded and sound attenuated room at the Neuroscience Laboratories (Center for Brain and Cognition, University Pompeu Fabra, Barcelona, Spain). During the experiment, participants sat in a comfortable armchair and were asked to watch a silent movie from a distance of around 80 cm from the screen and to ignore the vowels played through two loudspeakers (Gigaworks T20, Creative) positioned on the right and left side of the screen at an intensity of approximately 70 dB.

The experiments consisted of two blocks in which the same stimuli and procedures were used. The only difference between the two blocks was the somatosensory manipulation. In one block, participants had to hold a small plastic spoon over their tongue (the “spoon” condition, see Figure 1), while in the other block no spoon was placed over their tongue (the “no-spoon” condition). The placement of a spoon between the tongue and palate aimed at displacing the tongue from its natural position to a downwards position. The order of the blocks was counterbalanced across participants, with a break of approximately 10 minutes between blocks.
To control the placement of the spoon, participants were asked to extend their tongue, place the rounded part of the spoon (around 4 cm long and 2.5 cm wide, with a total spoon length: 12.5 cm) over the tongue, and move the tongue with the spoon inside the mouth. To avoid artifacts in the EEG signal with the spoon, participants were told to close their mouth softly and avoid muscle movements that could be elicited by pressing their teeth or lips together.

The stimuli were displayed using a multi-feature paradigm (Näätänen, Pakarinen, Rinne, & Takegata, 2004; Pakarinen et al., 2009, 2013), where every standard stimulus is followed by one of the different deviants. The standard stimulus in this experiment was the vowel /e/ and the deviants were the six vowels /a/, /i/, /o/, /u/, /ö/ and /y/. The probability of each deviant being played was 16.7%. The deviants were presented pseudo-randomly. Before repeating a given deviant, all other deviants were presented once and the same deviant was never presented twice in a row. At the beginning of each sequence and after each break the standard stimulus was played 5 times.

Each block (spoon or no-spoon) was approximately 12.5 minutes long. To avoid discomfort in holding the spoon for too long, the audio stopped every 2.5 minutes and the participants took a self-paced break. Recordings from the first 30 seconds after each break were deleted, because movement artifacts could occur due to the positioning of the spoon. During each spoon block each deviant was played 100 times, and the standard was played 605 times, because of the exclusion of the 30 initial seconds of the recording following a break, 25 additional presentations of each deviant and 175 additional presentations of the standard were not analyzed. The stimulus onset
asynchrony (SOA) was 488 ms. The experiment was programmed using the Psychtoolbox (Brainard, 1997; Pelli, 1997) running on MATLAB® 7.9.0.529 (R2009b, The MathWorks Inc., Natick, MA).

Electroencephalography recordings and analysis

ERPs were recorded using 31 electrodes mounted in an electrocap (Electro-Cap International). Additionally, five single electrodes were placed on the right and left mastoids (RM and LM), the infraorbital ridge and the outer canthus of the right eye (to measure vertical and horizontal eye movements), and the reference was placed on the tip of the nose. The electrode impedances were kept below 5 kOhm and the sampling rate was 500 Hz.

For the pre-processing, BrainVision Analyzer (v. 2.0; Brain Products) was used. The EEG was offline filtered between 0.1 and 30 Hz (slope: 12 dB/oct) and a notch filter was applied at 50 Hz. Eye movements were corrected using the ICA (independent component analysis) ocular correction. Data from one participant was excluded because of problems with the ICA ocular correction. Epochs with EEG exceeding either ± 200 μV at any channel (except the ones used for the ocular correction), activity < 0.5 μV, or voltage step/sampling > 50 μV within intervals of 200 ms were automatically rejected off-line. Visual inspection of the individual data after the ICA ocular correction and artifact rejection showed that the combination of both procedures reliably marked and corrected bad segments in all participants. The data from the experimental trials was segmented into epochs from -100 ms pre-stimulus to 500 ms post-stimulus onset.
The segmented data was analyzed with EEGLAB v12.0.2.5b (Delorme & Makeig, 2004) running on MATLAB® 8.2.0.701 (R2013b, The MathWorks Inc., Natick, MA). For each segment the baseline correction and DC Detrend were applied. The artifact-free epochs were averaged separately for each participant, spoon condition (spoon and no-spoon), standard, and deviant vowel group. Deviant vowel groups were created by combining vowels with the same vowel height and nativeness, giving the following groups: low/mid native (/a/, /o/), low/mid unknown (/ö/), high native (/i/ and /u/) and high unknown (/y/). The percentage of artifact-free epochs for each participant was higher than 80% (mean: 98.719% ± 0.464) and an ANOVA showed no differences for the percentage of artifact-free epochs for the block order (spoon tongue first, spoon tongue second: F(21,32) < 1, p = 0.955), spoon conditions (F(1,32) = 1.644, p = 0.209), the vowel type (standard and the four deviant vowel groups: F(4,128) < 1, p = 0.843), or any interaction between the factors (all p > 0.1).

MMNs were computed for each participant by subtracting the average ERPs of the standard from the average ERPs of each deviant vowel group. The peak amplitude of the MMN was identified in the difference grand averages as the most negative peak in the time window between 100 and 200 ms at electrode Fz for each deviant vowel group. The average peak amplitude was calculated for a 40 ms time window centered in the MMN maximum peak. The MMN amplitudes for two participants were excluded from the analysis because their MMN exceeded 2.5 standard deviations of the mean for the low unknown deviant, one in the spoon block and the other in the no-spoon block. Further statistical analyses were performed with R (R Core Team, 2014). The average MMN amplitudes were tested against zero using one sample t-tests (two-tailed) to test if each deviant vowel group (i.e. vowel height * nativeness) elicited a reliable MMN in
each spoon condition (spoon, no-spoon) and order (ST1: spoon tongue first, or ST2: spoon tongue second). For the analysis, a mixed ANOVA was conducted for the MMN peak amplitude including the between-subjects factor order (ST1, ST2) and the within-subjects factors spoon condition (spoon, no-spoon), vowel height (high, low/mid), and nativeness (native, unknown). In case of significant interactions, we performed t-tests between pairs of variables, correcting the p-values for multiple comparisons using the Bonferroni correction. Additionally, for the main findings we calculated the Bayes Factor (BF) by means of a Bayesian t-test using the Bayes Factor package in R. In a paired t-test, BF reports the likelihood ratio of the alternative hypothesis versus the null hypothesis. In general, BFs higher than 1 are considered to show a high likelihood for differences between conditions, above 3 are considered a substantial likelihood for differences between the conditions, and below 0.3 are considered substantially not different likelihoods between the conditions (Dienes, 2011).

**Results**

T-tests testing the MMNs against zero (Table 1) showed significant MMNs for all deviant groups and spoon conditions in at least one order, except for the low unknown vowel (/ö/) presented in the ST2 order. An ANOVA including the between-subjects factor order and the within-subject factors spoon condition, vowel height, and nativeness (Figure 2) showed a significant main effect of order (F(1,32) = 7.995, p = 0.008, BF t-test = 5.960) caused by larger MMNs for spoon over the tongue first (-1.678 µV ± 0.130) compared to spoon over the tongue second (-0.901 µV ± 0.122). The main effect for vowel height was also significant (F(1,32) = 83.528, p < 0.001, BF t-test = 25,265,748), the MMN was larger for high vowels (-1.956 µV ± 0.143) compared to low vowels (-0.623 µV ± 0.110). There was a marginal interaction between order and
vowel height (F(1,32) = 4.157, p = 0.050). T-tests comparisons were used to investigate the interaction further (Table 2). They showed that the two orders had similar MMNs for the low deviants (t(32) = -2.301, p = 0.168) and were marginally different for the high deviants (ST1-ST2: t(32) = -2.773, p = 0.055). In line with the main effect for vowel height, within the same order, the high deviants elicited larger MMNs than the low deviants (ST1 low – ST1 high: t(16) = 7.656, p < 0.001, ST2 low – ST2 high: t(16) = 5.195, p < 0.001). The high deviant for the ST1 order elicited a larger MMN than the low deviant for the ST2 order (t(32) = 6.843, p < 0.001). In contrast to general larger MMNs for the high vowels, the high deviant for the ST2 order elicited a similar MMN than the low deviant for the ST1 order (t(32) = 1.772, p = 0.515). No other effects reached significance. There was no effect of the spoon condition (F(1,32) = 0.017, p = 0.898, BF t-test = 0.185), nativeness (F(1,32) = 0.017, p = 0.898, BF t-test = 0.203), or any interactions including these variables (all p > 0.05).

Discussion
In this study we investigated if a somatosensory manipulation on a speech articulator—a spoon over the tongue—influences speech perception during passive listening, as measured by the MMN amplitude, of familiar and unfamiliar speech sounds. Based on previous experiments by Möttönen and colleagues (Möttönen et al., 2013, 2014) suggesting that in passive listening conditions the production of all articulators is modeled, we expected that the spoon over the tongue would affect the MMN amplitude independently of the articulation of the vowel (low vs. high tongue height) but only for native phonemes. Specifically, we formulated two hypotheses regarding the influence of the spoon. If the spoon over the tongue activates the tongue muscle, we would expect that this excitability would result in an increased MMN amplitude, similar to a previous
TMS study that showed improved behavioral performance for speech sounds involving the tongue after exciting the tongue representation in the motor cortex (D’Ausilio et al., 2009). Alternatively, the spoon could inhibit the simulation of the movements of the tongue, leading to a decrease in the MMN amplitude in line with previous rTMS studies showing that deactivating the lip representation in the motor cortex leads to decreased behavioral performance and MMN amplitudes when discriminating speech sounds (Möttönen et al., 2013; Möttönen & Watkins, 2009). For unknown phonemes, we expected similar MMN for the spoon and no-spoon condition based on the claim by Wilson and Iacoboni (2006) that the production system cannot model the gesture for unfamiliar speech sounds. Contrary to these hypotheses, results showed no main effect of the spoon, or interaction between the spoon manipulation and vowel familiarity, meaning that the spoon manipulation did not modulate the processing of either type of vowel (familiar or unfamiliar). Furthermore, the Bayes Factor (< 1) showed a very low likelihood of differences between the spoon and no-spoon conditions. Unexpectedly, the order of the experimental blocks, namely whether participants started with or without the spoon over the tongue, had a great impact on the MMN response. There was a global increase of the MMN amplitude when participants started with the spoon on the tongue, as indicated by the significant order effect and the Bayes Factor above 3 suggested a substantial likelihood for differences between the two orders.

For both orders, we expected a different MMN when no spoon was in the mouth, as compared to when the spoon was in the mouth. However, the MMN amplitude did not change when the spoon was removed or introduced (Figure 3). The order effect, together with the lack of a spoon effect, indicates that, rather than the presence of the spoon, the MMN was modulated by the experimental situation the participants faced.
when starting the experiment. One explanation for the present results is that the MMN reflected a combination of perceptual after-effects and response habituation. The similar MMN amplitude regardless of the presence of the spoon for the participants that started with the spoon in the mouth could reflect a perceptual after-effect whereas for the participants that started with the no-spoon block it could be caused by habituation to the auditory stimulus. Previous research has reported after-effects of somatosensory manipulations in speech perception and production (Leung & Ciocca, 2011; Nasir & Ostry, 2009). Nasir & Ostry (2009) trained participants to articulate words involving a large jaw movement, such as “had” and “bad”, while their jaw was being displaced with a robotic device. Before and after the motor training, the authors measured participants’ perceptual categorization of vowels varying in their degree of jaw movement on a continuum between the vowels “head”, articulated with a small jaw movement, and “had”, articulated with a large jaw movement. Results showed that the training distorted participants’ perceptual skills. After the motor training, when the robotic device had been removed, more words were perceived as articulated with a small jaw movement, the word “head”, as compared to before the training. The authors interpret the perceptual shift to be caused by an after-effect induced by the motor training. Thus, it is feasible that in the present study the spoon over the tongue had similar after-effects in the no-spoon block for the participants starting with the spoon. Regarding habituation effects, the MMN amplitude has been shown to decrease in time, showing a higher MMN amplitude in the first 5.5 minutes compared to the 22-27.5 minutes of the recording (McGee et al., 2001). In our experiment we suggest that for the participants starting with the no-spoon block, habituation of the MMN amplitude in time interfered with the expected increase of the MMN amplitude due to the spoon manipulation. This combination of after-effects and habituation fits with the present findings, showing a
general enhancement for the participants starting with the spoon over the tongue and no changes for either spoon order over time.

Furthermore, high vowels elicited a larger MMN than the low ones. Yet, when participants started the study with the spoon over the tongue, the low vowels elicited a MMN comparable to those of high vowels when participants started with no-spoon. That is, when beginning the study with the spoon over the tongue, low vowels evoked a MMN as large as high vowels in the reverse block order. This interaction between the vowel height and order gives some indications for the spoon order influencing speech perception in an articulatory specific manner.

An alternative explanation to the order effect presupposes that starting with holding the spoon over the tongue might result in overall increased attention, affecting the processing of the auditory stimuli. Indeed, although the MMN can be reliably elicited in passive listening conditions, attending to the stimuli results in larger MMN amplitudes (Deguchi et al., 2010; Erlbeck, Kübler, Kotchoubey, & Veser, 2014; Sussman, Ritter, & Vaughan, 1998; Sussman, Winkler, & Wang, 2003; Sussman, 2013; Szymanski, Yund, & Woods, 1999; Tervaniemi et al., 2009; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). Participants starting with the spoon over the tongue may be more attentive to the auditory stimulation throughout the experimental situation, even when the spoon was removed. In contrast, participants running the spoon condition second were already used to the experimental situation and the stimuli when the spoon was introduced and, hence, no increase in attention level was evoked in these participants by the spoon over the tongue. The increased attention when the participants started with the spoon over the tongue, as compared to starting
without the spoon, might be the cause for the observed global increase of the MMN amplitude in this condition. Summing up, the global increased of MMN amplitude when participants started with the spoon over the tongue could have two potential sources, either the manipulation of a speech articulator which had a long-lasting after-effect, or an increase in attention paid to the stimulus when the experimental situation required a motor task from the beginning. To test these two alternatives, we ran Experiment 2.

**Experiment 2**

In Experiment 2 the same stimuli and procedures as in Experiment 1 were used with the exception that participants performed a motor task that did not involve any speech articulator, i.e., holding a spoon in the hand, and all participants started with the motor task. The spoon in the hand should engage the same attentional demands as holding the spoon over the tongue albeit not tapping onto a speech articulator muscle. We expected that if the larger MMN amplitude for the participants starting with the spoon over the tongue was due to an after-effect of stimulating an articulator muscle, then starting with the spoon in the hand should result in a lower MMN amplitude compared to starting with the spoon over the tongue. Alternatively, if the order effect was due to higher attention because of holding a spoon, the MMN amplitude for the participants starting with the spoon in the hand should be similar to the participants starting with the spoon over the tongue.

**Methods**

**Participants**

Eighteen Spanish listeners (21.33 ± 3.48 years, 15 females) recruited from the same population as in Experiment 1 were tested. Data from one participant was excluded in
an outlier analysis because the MMN amplitude was more than 2.5 standard deviations from the mean in two of the MMN conditions, leaving a final sample of seventeen participants.

Stimuli and procedure

Stimuli and procedure were the same as in Experiment 1, except that participants now held a spoon in the hand (Figure 1). Participants held the spoon between their fingers of their right hand. The arm lay relaxed on the armrest with the hand outside the armchair. All participants started with the spoon in the hand, followed by the no-spoon block.

Electroencephalography recordings and analysis

Electroencephalography recordings were done following exactly the same procedures as in Experiment 1. The percentage of artifact-free epochs for each participant was higher than 80% (99.370% ± 0.272) and an ANOVA including both Experiment 1 and 2 showed no differences for the percentage of artifact-free epochs for the experiment type (Experiment 1 spoon tongue first, Experiment 1 spoon tongue second, Experiment 2 spoon hand first: F(2,48) < 1, p = 0.645), the spoon conditions (F(1,48) = 1.353, p = 0.250), the vowel type (standard and the four deviant vowel groups: F(4,192) = 1.063, p = 0.376), or any interaction between the factors (all p > 0.1). The MMN amplitudes for one participant were excluded from the analysis because they exceeded 2.5 standard deviations of the mean for high unknown spoon and no-spoon MMN.

Data analysis was the same as in Experiment 1. First, the average MMN amplitudes were tested against zero using one sample t-tests (two-tailed) to test if each deviant vowel group (i.e. vowel height * nativeness) elicited a reliable MMN in each spoon
condition (spoon, no-spoon). Then, a mixed ANOVA was conducted, pooling the data from both Experiments 1 and 2. The ANOVA included the between-subject factor experiment type (ST1: spoon tongue first, ST2: spoon tongue second, SH1: spoon hand first) and the within-subject factors spoon condition (spoon, no-spoon), vowel height (high, low/mid), and nativeness (native, unknown). Additionally, for the main findings we calculated the Bayes Factor (BF) by means of a bayesian t-test.

**Results**

T-tests showed reliable MMNs (Table 1) for all MMNs. An ANOVA including the between subjects factor experiment type and the within subject factors spoon condition, vowel height and nativeness showed a significant main effect of experiment type (F(2,48) = 5.954, p = 0.005; Figure 2) caused by significant larger MMNs for spoon over the tongue first and spoon in the hand first compared to spoon over the tongue second (ST1 - ST2: t(32) = 2.818, p = 0.008, BF t-test = 5.960, SH1 - ST2: t(32) = 3.124, p = 0.004, BF t-test = 10.712; ST1: -1.678 µV ± 0.130, SH1: -1.860 µV ± 0.130, ST2: -0.901 µV ± 0.122). No differences were found between spoon over the tongue first and spoon in the hand first (ST1 - SH1: t(32) = 0.601, p = 0.552, BF t-test = 0.378). As in Experiment 1, there was a significant main effect of vowel height (F(1,48) = 76.327, p < 0.001, BF t-test = 329,915,628) caused by a larger MMN to high vowels (-2.081 µV ± 0.110) compared to low vowels (-0.878 µV ±0.087). No other effects reached significance. There was no effect of the spoon condition (F(1,48) = 0.336, p = 0.565, BF t-test = 0.178), nativeness (F(1,48) = 2.000, p = 0.164, BF t-test = 0.391), or any interaction including these variables (all p > 0.05).

**Discussion**

23
The goal of Experiment 2 was to investigate the cause for the order effect found in Experiment 1. Specifically, in Experiment 1 we found that starting with the spoon over the tongue resulted in higher MMN compared to the spoon over the tongue second condition. In Experiment 2 we tested if the increase in MMN can also be found for a somatosensory manipulation on a non-articulator—a spoon in the hand. With this experiment we aimed at differentiating between two possible explanations for the order effect. First, starting with the spoon over the tongue might have led to articulatory specific after-effects lasting into the subsequent no-spoon condition. If this possibility were true, we would have expected that starting with the spoon over the tongue results in a larger MMN compared to starting with the spoon in the hand. The second possibility was that focusing on holding the spoon over the tongue might lead to higher attention for the participants starting with a somatosensory manipulation—individually of the manipulation involving an articulator or not. Results of the second experiment are in line with the second explanation. They showed that the MMN was similar when a somatosensory manipulation was introduced at the beginning of the study, regardless of the muscles manipulated (i.e., tongue or hand) and increased when starting with the manipulation, as compared to starting with none. Bayes Factors above 3 revealed a substantial likelihood for differences when starting with the spoon (either in the hand or mouth) and starting without the spoon. The increased MMN in Experiment 2 cannot have a speech origin, since no speech articulator was manipulated. The similar MMN enhancement for the spoon over the tongue first (Experiment 1) and the spoon in the hand first (Experiment 2), compared to the spoon over the tongue second (Experiment 1), indicates a general attentional effect of starting with a somatosensory manipulation. The attentional effect of the MMN found in Experiments 1 and 2 is in line with a variety of studies that show that although the MMN can be
elicited during passive listening conditions, attention modulates the MMN amplitude (Deguchi et al., 2010; Erlbeck, Kübler, Kotchoubey, & Veser, 2014; Sussman, Ritter, & Vaughan, 1998; Sussman, Winkler, & Wang, 2003; Sussman, 2013; Szymanski, Yund, & Woods, 1999; Tervaniemi et al., 2009; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998).

Previous studies showed no influence of non-articulators (hand) on speech perception (Möttönen et al., 2013; Möttönen & Watkins, 2009). In these studies, the authors deactivated either the lip or hand representation in the primary motor cortex by means of rTMS and then recorded speech perception responses, behavioral identification, and discrimination of stimuli ranging from a continuum between lip and tongue articulated phonemes, or the MMN for changes between lip and tongue articulated phonemes. In all cases, only the deactivation of the lip representation, but not the hand representation changed behavioral performance or the MMN amplitude. In those studies articulators were manipulated before the experimental trials started. In contrast, in our study participants had to monitor the somatosensory manipulation during the experimental trials—taking care that the spoon did not fall. Additionally, introducing the spoon at the beginning of the study might intrigue participants about the function of the spoon in the study and make them more alert. The task demands and participants’ expectations are likely to explain the enhanced attention found for the participants starting with the somatosensory manipulation in our study.

**General discussion**

The goal of the present study was to investigate the specificity of the speech production system in its contribution to speech perception in passive listening conditions. To this
end we tested if a somatosensory manipulation on a speech articulator—placing a spoon over the tongue—influences the MMN amplitude of familiar and unfamiliar speech sounds varying in vowel height. Based on previous experiments by Möttönen and colleagues (Möttönen et al., 2013, 2014) showing that the production of all articulators is modeled in passive listening conditions, we expected that the spoon over the tongue affects the MMN amplitude independently of the somatosensory manipulation effecting the articulation of the vowel (low vs. high tongue height) for familiar phonemes. We formulated two contrasting hypotheses regarding the effect of the spoon over the tongue for native vowels. The spoon over the tongue might activate the tongue muscle thereby increasing the MMN amplitude in line with a previous TMS study showing that increasing the excitability of the tongue representation in the motor cortex leads to increased behavioral performance involving phonemes articulated with the tongue (D’Ausilio et al., 2009). On the contrary, the spoon over the tongue might inhibit the ability to simulate the tongue movement and thus reduce the MMN amplitude. This second hypothesis is in line with previous rTMS studies showing that deactivating the lip representation in the motor cortex led to reduced behavioral performance and a reduced MMN amplitude when discriminating phoneme contrasts (Möttönen et al., 2013; Möttönen & Watkins, 2009). For unfamiliar phonemes we did not expect an influence of the spoon on the MMN amplitude following the claim that the production system cannot model the gesture for unfamiliar speech sounds (Wilson & Iacoboni, 2006).

In Experiment 1, we did not find any difference in the MMN amplitude between the spoon conditions, i.e., whether the spoon was over the tongue, or between native and unknown vowels. Instead, we found that starting with a somatosensory manipulation, a
spoon over the tongue, resulted in a larger MMN throughout the experiment. We hypothesized that the order effect might be due to a combination of after-effects and habituation. For the participants starting with the spoon, we propose that the MMN amplitude did not decrease in the no-spoon block due to an after-effect. For the participants starting with the no-spoon block, the MMN amplitude did not increase when the spoon was introduced due to habituation to the stimulus. After-effects induced by somatosensory manipulations have been reported previously for speech production and speech perception (Leung & Ciocca, 2011; Nasir & Ostry, 2009) and response habituation has been reported to show a decrease of the MMN amplitude in time, especially during the first 20 minutes of the recording (McGee et al., 2001). An alternative to this after-effect and habituation explanation is that it is also possible that the participants starting with the spoon over the tongue were more alert due to the motor task, or they were more intrigued about the spoon manipulation and therefore paid more attention to the task and stimuli resulting in a higher MMN amplitude throughout the experiment. Although the MMN can be elicited in passive listening conditions, active MMN tasks have shown increases of the MMN amplitude (Deguchi et al., 2010; Erlbeck, Kübler, Kotchoubey, & Veser, 2014; Sussman, Ritter, & Vaughan, 1998; Sussman, Winkler, & Wang, 2003; Sussman, 2013; Szymanski, Yund, & Woods, 1999; Tervaniemi et al., 2009; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998). In this case, the MMN amplitude for the participant starting with the no-spoon condition, attention is lower. When the spoon is introduced in the second block, participants are already used to the task and stimuli, resulting in no effect of the spoon.
Experiment 1 also showed an interaction between the order of the spoon conditions (starting with the spoon over the tongue or the spoon condition second) and the vowel height. This interaction showed that, when starting with the spoon, low vowels elicited a MMN as large as high vowels in the reverse block order, despite low vowels globally evoking a smaller MMN response than high vowels in all other experimental conditions (as indicated by the main effect of vowel height). This interaction suggested that the order could have a specific articulator effect on speech processing rather than an attentional effect.

To test the different possibilities for the order effect found in Experiment 1—the after-effect explanation or the general attention explanation—we ran Experiment 2. The experiment was the same as Experiment 1 with the exception that participant started with a somatosenory manipulation on a non-articulator—a spoon in the hand. Holding a spoon in the hand should result in similar attentional effects than in Experiment 1 but without engaging an articulatory muscle. We expected that if the order effect was due to articulatory specific after-effects lasting into the no-spoon condition, starting with a spoon in the hand should result in similar amplitudes as starting with the no-spoon condition and a lower amplitude than starting with the spoon over the tongue. Alternatively, if the order effect was caused by general higher attention for the participants starting with the somatosensory manipulation, starting with the spoon in the hand should result in comparable amplitudes as starting with the spoon over the tongue. Results were in line with the second explanation, showing a similar MMN amplitude for participants starting with the spoon in the hand and over the tongue.
Experiments 1 and 2 showed similar MMN enhancements when the participants started the experiment with the spoon over their tongue (Experiment 1) or in the hand (Experiment 2). This effect cannot be caused by motor contributions to speech perception, given that the enhancement was also present for a non speech articulator, the hand. The MMN amplitude can also be modulated by the physical properties of the stimuli. However, in the present study, the same speech material was present for all the experimental conditions (with and without the spoon in Experiments 1 and 2). Rather, the increase of the MMN amplitude was related to the time when the spoon was introduced in the experiment and was independent of the muscle involved in the somatosensory manipulation. Therefore, the increased MMN seems to be consistent with a general attentional or alertness effect due to starting the experiment with an experimental manipulation, i.e., the spoon. We propose that participants were more attentive to the stimuli when starting with the spoon as they might be more intrigued about the relation between the spoon and the sounds or simply because the motor task (holding the spoon without letting it fall) might have increased alertness. In comparison, the participants with the spoon condition second were less alert in the first no-spoon block and were already used to the experimental situation and the sounds when the spoon was introduced in the second block. The attentional effect of the MMN found in this experiment is in line with a variety of studies showing that attention modulates the MMN amplitude (Deguchi et al., 2010; Erlbeck, Kübler, Kotchoubey, & Veser, 2014; Sussman, Ritter, & Vaughan, 1998; Sussman, Winkler, & Wang, 2003; Sussman, 2013; Szymanski, Yund, & Woods, 1999; Tervaniemi et al., 2009; Woldorff, Hackley, & Hillyard, 1991; Woldorff, Hillyard, Gallen, Hampson, & Bloom, 1998).
The absence of an effect of the spoon over the tongue manipulation on speech perception contrasts with previous studies showing influences of manipulating the speech production system on speech perception when applying TMS (D’Ausilio et al., 2009; Möttönen & Watkins, 2009) or somatosensory manipulations (Ito et al., 2009; Nasir & Ostry, 2009). The somatosensory manipulations used in previous studies involved stretching the skin next to the lips upwards or downwards (Ito et al., 2009) or learning how to articulate words when the jaw was being displaced (Nasir & Ostry, 2009). While TMS studies investigated the perception of consonants, the somatosensory studies investigated the perception of vowels from a continuum between the words “head” and “had” (i.e., /ɛ/ vs. /ɛ/). Consonants and vowels are acoustic and articulatory very different: consonants are fast-transitions articulated with a closure of the vocal tract and vowels are steady-state sounds articulated with an open vocal tract. Despite the distinct methodologies and speech stimuli employed by previous studies, they consistently found an effect of manipulating the speech production system on speech perception. While the present study also investigated the perception of vowels, it was methodological very different from previous somatosensory studies: previous studies manipulated a different articulator (i.e., skin next to the lips or jaw as compared to the tongue in the present study) and the somatosensory manipulation involved a movement while in our study the spoon was continuously pressing the tongue. The somatosensory manipulations applied by Ito et al. (2009) and Nasir and Ostry (2009) might thus have affected speech perception to a greater degree than the spoon over the tongue used in our study. The repetition of the movement in these studies might have activated the motor cortex constantly and, possibly, in a stronger way, than the continuous pressure on the tongue to which the motor cortex might have adapted over time. Furthermore, Ito et al. (2009) and Nasir and Ostry (2009) used a speech perception behavioral task which
involved attention, while in our study we measured speech perception in passive
listening conditions. As explained before, attention has been shown to modulate the
specificity of motor influences in speech perception (Möttönen et al., 2013, 2014) and
the attentive tasks used in the previous studies are likely to have resulted in stronger
effects of the somatosensory manipulation on speech perception. Thus, the type of
sensorymotor manipulation and the amount of attention allocated on the speech stimuli
likely influence the extent to what the speech perception is disturbed by a
somatosensory manipulation.

A lack of an effect of the spoon manipulation on speech perception might reflect the
robustness of the mature speech perception system rather than refuting a role of the
speech production system during perception. In a previous study with preverbal infants
a similar somatosensory manipulation to the present one, i.e., blocking the tongue with a
flat teething toy, impaired speech discrimination of a non-native, unknown (Hindi)
speech contrast. Thus, manipulating the tongue position influenced speech perception
even in the absence of productive or perceptive experience with the speech sounds. The
contrasting findings in preverbal infants and in adults in the present study suggest that
somatosensory influences are stronger during early stages of development. Two
alternative explanations can account for the distinct results of the somatosensory
manipulation in infants and adults. The involvement of the speech production system
may play a more fundamental role during initial stages of speech learning than during
adulthood. Yet, this interpretation is not supported by previous studies with adult
populations that showed poor speech perception when the motor system was disrupted
(D’Ausilio et al., 2009; Möttönen & Watkins, 2009, Ito et al., 2009; Nasir & Ostry,
2009). Alternatively, the adult speech perception system may be less sensitive to
somatosensory disturbances. In line with the former interpretation, Bruderer et al. (2015) discussed the existence of a sensorimotor–articulatory perturbation threshold for accurate speech perception. Such a threshold may be different depending on the abilities of the learners and, hence, a similar sensorimotor perturbation may affect speech perception to a greater extent in inexperienced speech listeners. We suggest that during development the perception system may not only learn to accurately perceive speech but also to cope with auditory (i.e., noise, Newman, 2005) and sensorimotor perturbations presented in every-day life situations to achieve a robust perception. Thus, through learning, the sensorimotor–articulatory perturbation threshold for accurate speech perception may be less restrictive. Future studies may reveal whether the sensitivity of the speech perception system to sensorimotor disturbances changes across the lifespan.

Previous studies have shown influences of the production system in speech perception (D’Ausilio et al., 2009; Ito et al., 2009; Möttönen et al., 2013, 2014; Möttönen & Watkins, 2009; Nasir & Ostry, 2009; Sato et al., 2011), however, whether the production system contributes to speech perception in easy tasks and without attentional demands is highly debated (Hickok et al., 2011; Scott et al., 2009). Yet, the current results cannot add to this debate and rather show that the MMN can be influenced by the alertness or attention of the participants when an irrelevant task for speech perception, holding a spoon over the hand or the tongue, is performed at the beginning of the study. These results suggest that MMN studies that combine passive listening with somatosensory manipulations may be capturing attention processes when assessing the MMN.


Tables

Table 1: Experiment 1 & 2. T-tests of the MMN mean amplitude centered at a 40ms time window at the peak latency at Fz for the different orders in the two experiments (ST1: spoon tongue first, ST2: spoon tongue second, SH1: spoon hand first), spoon conditions, vowel height and nativeness.

<table>
<thead>
<tr>
<th>Deviants</th>
<th>Order</th>
<th>Latency (ms)</th>
<th>Amplitude (µV)</th>
<th>Latency (ms)</th>
<th>Amplitude (µV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-spoon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Native</td>
<td>ST1</td>
<td>129</td>
<td>-1.182*</td>
<td>129</td>
<td>-0.482^</td>
</tr>
<tr>
<td>Low Unknown</td>
<td>ST1</td>
<td>183</td>
<td>-0.877*</td>
<td>167</td>
<td>-0.909*</td>
</tr>
<tr>
<td>High Native</td>
<td>ST1</td>
<td>121</td>
<td>-2.632*</td>
<td>131</td>
<td>-2.294*</td>
</tr>
<tr>
<td>High Unknown</td>
<td>ST1</td>
<td>123</td>
<td>-2.612*</td>
<td>133</td>
<td>-2.435*</td>
</tr>
<tr>
<td>Low Native</td>
<td>ST2</td>
<td>113</td>
<td>-0.290</td>
<td>111</td>
<td>-0.534*</td>
</tr>
<tr>
<td>Low Unknown</td>
<td>ST2</td>
<td>181</td>
<td>-0.152</td>
<td>191</td>
<td>-0.555</td>
</tr>
<tr>
<td>High Native</td>
<td>ST2</td>
<td>113</td>
<td>-1.200*</td>
<td>113</td>
<td>-1.441*</td>
</tr>
<tr>
<td>High Unknown</td>
<td>ST2</td>
<td>135</td>
<td>-1.456*</td>
<td>139</td>
<td>-1.578*</td>
</tr>
<tr>
<td>Spoon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Native</td>
<td>SH1</td>
<td>137</td>
<td>-1.190*</td>
<td>143</td>
<td>-1.282*</td>
</tr>
<tr>
<td>Low Unknown</td>
<td>SH1</td>
<td>177</td>
<td>-1.473*</td>
<td>159</td>
<td>-1.610*</td>
</tr>
<tr>
<td>High Native</td>
<td>SH1</td>
<td>127</td>
<td>-2.059*</td>
<td>145</td>
<td>-2.159*</td>
</tr>
<tr>
<td>High Unknown</td>
<td>SH1</td>
<td>149</td>
<td>-2.180*</td>
<td>147</td>
<td>-2.928*</td>
</tr>
</tbody>
</table>

Significant differences: ^ p<0.1, * p<0.05.
Table 2: Experiment 1. T-tests for the interaction between order and vowel height.

<table>
<thead>
<tr>
<th>comparisons</th>
<th>t-test type</th>
<th>degrees of freedom</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST1 low ST1 high</td>
<td>paired</td>
<td>16</td>
<td>7.656</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>ST2 low ST2 high</td>
<td>paired</td>
<td>16</td>
<td>5.195</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>ST1 low ST2 low</td>
<td>two-samples</td>
<td>32</td>
<td>-2.301</td>
<td>0.168</td>
</tr>
<tr>
<td>ST1 high ST2 high</td>
<td>two-samples</td>
<td>32</td>
<td>-2.773</td>
<td>0.055^</td>
</tr>
<tr>
<td>ST1 low ST2 high</td>
<td>two-samples</td>
<td>32</td>
<td>1.772</td>
<td>0.515</td>
</tr>
<tr>
<td>ST2 low ST1 high</td>
<td>two-samples</td>
<td>32</td>
<td>6.843</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

Significant differences: ^ p<0.1, * p<0.05.
Figure 1. Methods. Right: First and second formant frequency values in Hz for the Spanish vowels /a/, /e/, /i/, /o/, /u/ and the Finnish vowels /ö/ and /y/. Left: Placement of the spoon over the tongue (up, experiment 1) and in the hand (down, experiment 2).
Figure 2. Experiment 1 & 2: Grand-average MMN at Fz for the vowels grouped by vowel height (down: low, up: high) and nativeness (left: native, right: unknown). In each plot, solid lines depict the difference waves for spoon over the tongue first (ST1), light dashed lines depict the difference waves for those participants starting with no spoon over the tongue second (ST2), and thin lines depict the difference waves for spoon in the hand first (SH1). Red lines display the difference waves for the spoon blocks and blue lines display the difference waves for the no-spoon blocks.
Figure 3. Experiment 1: MMN amplitude measured at Fz for the different orders and blocks (1\textsuperscript{st} or 2\textsuperscript{nd}). Solid lines depict the order spoon over the tongue first (ST1) and dashed lines depict the order spoon over the tongue second (ST2). The figure shows that the MMN amplitude stays the same within the same order and does not increase or decrease in time.