Allophonic familiarity differentiates word representations in the brain of native speakers of regional linguistic varieties.

Giuseppe Di Dona[1], Federica Mantione[1], Birgit Alber[2], Simone Sulpizio[3,4], & Francesco Vespignani[5]

[1] Dipartimento di Psicologia e Scienze Cognitive, Università degli Studi di Trento, Italy

[2] Facoltà di Scienze della Formazione, Libera Università di Bolzano, Italy

[3] Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Italy

[4] Milan Center for Neuroscience (NeuroMi), University of Milano-Bicocca, Italy

[5] Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università degli Studi di Padova, Italy

Author Note

Correspondence concerning the article should be addressed to: Giuseppe Di Dona, Dipartimento di Psicologia e Scienze Cognitive, Università degli Studi di Trento, Corso Bettini 84, 38068 – Rovereto (TN), Italy. e-mail: giuseppe.didona@gmail.com

Abstract

This study aims to shed light on the issue whether familiar allophonic variation is encoded in word representations. Both Italian speakers born in Trentino and speakers born in the Central-Southern regions of Italy took part in the experiment. We tested the MMN elicited by the same word encompassing two different allophones, one of which was more familiar to one group of participants than to the other, depending on their regional variety of Italian. The Trentino group showed an enhanced MMN for the word embedding the familiar variant while Central-Southern speakers showed no difference. The amplitude of the MMN for the unfamiliar word variant in Trentino speakers showed an inverse correlation with the passive exposure to the Trentino dialect. We conclude that words embedding familiar and unfamiliar allophones are differently represented in the brain of native speakers of regional language and the degree of differentiation is modulated by individual experience.

Keywords**:** Word representation; Multilingualism; Regional language; Allophonic variation; Allophonic familiarity; EEG; Mismatch Negativity; MMN; Memory retrieval; Individual differences

1. Introduction

Listeners can recognize words despite the great amount of physical variability in speech signals. This is possible thanks to complex cognitive mechanisms that allow robust speech perception in non-ideal conditions characterized by noise, ambiguity, and acoustic-phonetic variations (Eisner & McQueen, 2018). This last source of variability may be particularly detrimental to speech perception as in order to understand words, acoustic-phonetic features must be mapped onto defined sound units (Liberman et al., 1967) that, when combined, form meaningful words. Phonemes are thought to be the fundamental sound units that, at the word level, cannot be interchanged without altering or disrupting the meaning (Trubetzkoy, 1969). For instance, the English word *bus* [bʌs] becomes *buzz* [bʌz] if the final /s/ phoneme is substituted with a /z/ phoneme, with [bʌs] and [bʌz] thus forming a minimal pair. Phonemes must be though as classes of sounds rather than single units with specific phonetic features. The members of such classes are defined as allophones, and their phonetic realization might change on the basis of specific rules. For instance, the initial and final /t/ phonemes in the English word *test*, are not phonetically equivalent: while the /t/ phoneme is produced as the aspirated [t^h] allophone when it is syllable initial, it is produced as [t] when it is syllable final (Ladefoged & Johnson, 2014). Conversely to phonemes, if two allophones of the same phonemic class (e.g., [t^h] and [t]) are interchanged, word meaning is not disrupted (Avery et al., 2008).

Linguistic productions are rich in allophonic variations (Chambers et al., 2002), which lead to different acoustic realizations of a same word. Systematic patterns of allophonic variation are typical to linguistic varieties (Fasold & Connor-Linton, 2014) and can be considered a solid cue to their identification (van Bezooijen & Gooskens, 1999). Speakers of different linguistic varieties of the same standard language can be familiar with different allophones that appear in specific phonological contexts and with specific frequency distributions within their respective varieties

(Calabrese, 2012; Chambers et al., 2002). While the cognitive (McClelland & Elman, 1986; Mitterer et al., 2013; Norris & McQueen, 2008) and neurophysiological mechanisms (Hickok & Poeppel, 2007) that map phonemes and other sub-lexical units onto lexical representations have been described in quite some detail, it is currently unclear: a) how the speech recognition system takes care of allophonic variation and still correctly recognizes words; b) to what extent the exposure to specific allophones, qualified as allophonic familiarity, can refine word representations and improve the recognition capacity. The present study aims to fill this gap by investigating the impact of allophonic variation on the retrieval processes of word representations. In so doing, this study also addresses the role of allophonic familiarity contingent to the quantity of exposure to a specific language variety.

Few studies so far used electrophysiological measures to investigate allophonic variations and allophonic familiarity. These studies focused on the Mismatch Negativity (MMN), (Näätänen et al., 2007), an ERP component thought to index specific memory retrieval processes. MMN is elicited when a change in auditory stimulation is detected, irrespectively of the listener's attention (Näätänen & Michie, 1979). MMN is typically measured using a passive-oddball task in which a sound is repeatedly presented (standard stimulus) and infrequently replaced by a different sound (deviant stimulus). Compared to the standard stimulus, the deviant stimulus elicits a larger negativity (i.e., the MMN) peaking ~150-250 ms after stimulus onset, and mainly visible on fronto-central electrodes (Näätänen, 1995). This effect is thought to index a violation of the representation of the standard sound in short-term memory (Näätänen et al., 2005). Moreover, to the best of our knowledge, no study has investigated allophonic discrimination and allophonic familiarity within meaningful words. These two aspects could be pivotal to explain how the speech recognition systems adapts to phonetic variability while correctly understanding words. Of particular interest with respect to our purpose is the finding that the MMN shows larger amplitude waveforms when the deviant stimulus is a phoneme or a word of the listener's native language, compared to when it belongs to an unknown language (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Cheour et al., 1998; Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2004). This *enhancement effect* has been interpreted as an index of a memory-trace retrieval process of phonemes and words from long-term memory. Therefore, the use of such a measure could be very well suited to reach the aim of our experiment, allowing us to implicitly test for the presence of specific memory traces for words embedding familiar allophones.

The outcomes of this study, which capitalizes on an innovative use of the MMN (i.e., an implicit electrophysiological measure) to investigate allophonic processing in real words, might be very informative for the theoretical accounts of spoken word recognition. We considered three main theoretical accounts of spoken word recognition — in which the impact of allophonic variation is addressed — to orient our experimental hypotheses. First, according to the inference-based account, speech is encoded into abstract features that activate single abstract lexical representations. In this class of models, pronunciation variants are accommodated to match the activated representation on the basis of inferential processes (Gaskell & Marslen-Wilson, 1998). This means that listeners exploit specific sets of rules to make inferences about the phonological viability of specific allophones before accessing the lexicon and retrieving word representations efficiently. Gaskell and Marslen-Wilson (1996) studied the place assimilation, a particular kind of phonetic variation that occurs at word boundaries. In English, the final consonant of a word can be articulated using the same place of articulation of the initial consonant of the following word. For instance, when word final coronals (e.g., /t/, /d/, /n/) are followed by word initial labials (e.g., /p/, /b/, /m/) the first are realized assimilating the place of articulation of the latter. The use of cross-modal repetition priming showed that when phonological variation results from an illegal assimilation of place of articulation (e.g., “wickib game”), word recognition is slower with respect to when phonological variation adheres to assimilation rules (e.g., “wickib prank”). Additionally, the effect was larger for meaningful words with respect to non-words, suggesting that the lexical status can facilitate the application of such assimilation rules. Further, the efficiency of inferential processes is thought to depend on the allophonic variant distribution within a language, meaning that large sets of similarly structured variants can lead to the generalization of the inferential processes (Pierrehumbert, 2006). Second, episodic models (Goldinger, 1998), instead, postulate that variants are integrated into lexical representations, meaning that surface phonetic details are always retained in the lexicon. In this scenario, successful word recognition depends on the degree of similarity between the input phonetic word form and the previously encountered variants, without the intervention of inferential cognitive processes that constrain lexical access. In a series of experiments, Goldinger (1996) showed that during word recognition performed on words uttered by different speakers, listeners performed with higher accuracy if in a previous exposure phase they heard the exact same word uttered by the same speaker, with respect to when the speaker was new. This suggests that listeners do store phonetic details of the words they hear, a conclusion which is in neat opposition with the notion of abstract word representations postulated in the inference-based account. Third, the hybrid approach theorizes the existence of multiple abstract representations of single words, and their activation is biased by the frequency of occurrence of an input variant, or in other words, allophonic familiarity (Connine & Pinnow, 2006). To this regard Pinnow & Connine (2014) studied schwa vowel deletion by which the English word *catholic* (/kæθəljk/) can be produced as /kæθljk/, without the schwa in the second syllable. Authors showed that in a lexical decision task with words embedding schwa deletions, words in which schwa deletion was more frequent were recognized faster with respect to words in which schwa deletion was less frequent. Further, exposure to low-frequency words for which schwa deletion had low frequency speeded up the recognition process in participants.

Additionally, in order to study the relationship between allophonic variation and allophonic familiarity during word recognition, it is necessary to consider how allophones are perceived in isolation first, as well as the influence of phonological contexts. Pallier, Bosch, and Sebastián-

Gallés (1997) investigated the discrimination of the [ɛ–e] contrast, which is allophonic in Spanish (i.e., [e] appears in closed syllables before m, n, t, θ, s as in [ba̲ˈlenθja̲], *Valencia* while [ɛ] appears in closed syllables when it is not followed by m, n, t, θ, s as in [ˈbɛlɣa̲], *Belgian*, but phonemic in Catalan (e.g., [te], *take* and [tɛ], *tea).* In a 2-AFC phoneme categorization task with stimuli from a [ɛ–e] continuum, participants were asked to report if the isolated vowels sounded more like the /e/ in the Catalan word for "*Pere*" ([perə], Peter), or more like the /ɛ/ in Catalan word "*pera*" ([pɛrə], pear). Secondly, they were asked to perform a same-different discrimination task with stimulus pairs from the continuum. Results showed clear categorical perception (i.e., steep categorization curve and high discrimination accuracy also with acoustically ambiguous stimulus pairs) in Spanish-Catalan bilinguals with Catalan-speaking parents with respect to Spanish Catalan bilinguals with Spanish-speaking parents. This result is in line with another study in which English and Spanish participants were asked to judge the similarity between different VCV sequences (e.g., ada, aɾa, aða) embedding either [d], [ð] or [ɾ]: both the [d–ð] and the [d–ɾ] contrast (the former being phonemic in English but allophonic in Spanish, the latter being the opposite) are rated as more similar when they are recognized as allophones with respect to when they are intended as phonemes (Boomershine et al., 2008). These two studies suggest that when allophonic contrasts are presented as isolated stimuli, they are harder to discriminate both implicitly (i.e., through categorization and discrimination tasks) and explicitly (i.e., through similarity judgements).

Although, contrastive results were found in a study by Peperkamp, Pettinato, and Dupoux (2003) in which French listeners performed a same-different task hearing the allophones [ʁ] and [χ] and phonemes /m/ and /n/ spliced into isolated VC syllables (e.g., aʁ, iχ; am, in). In another condition, participants performed the same task on VCCV sequences in which the first VC sequence was the test syllable embedding the allophone or the phoneme while the second CV syllable was the "context" syllable (e.g., as in aʁdi). Results showed that, when presented in isolation, phonemic and allophonic contrasts were both well discriminated by listeners, but when embedded in the VCCV sequences, allophonic contrasts were much more difficult to perceive with respect to phonemic contrasts and to allophonic contrasts presented in isolation. Further, authors showed that participants' allophonic discrimination accuracy improved after additional exposure task in which they listened to a list of VCCV where the first VC syllable embedded a stimulus of a [ʁ]-[χ] continuum. Authors suggested that despite the inconsistency of allophonic discrimination accuracy between their and previous studies, recorded when allophones are presented in isolation, the presence of a phonological context (i.e., the context syllable) largely hinders allophonic discrimination. However, this initial impairment in allophonic discrimination was reduced as a result of exposure.

Further, in a behavioural and EEG study on allophonic discrimination of isolated segments, Miglietta, Grimaldi, and Calabrese (2013) tested the discrimination of the [ɛ-e] and the /e/-/i/ contrast, respectively allophonic and phonemic for the speakers of Southern-Italian dialect of Tricase, spoken in Southern Apulia region. In line with Peperkamp et al. (2003), in a same-different discrimination task, participants could easily distinguish both the phonemic and the allophonic contrast. Additionally, as the MMN elicited by the /e/-/i/ phonemic contrast peaked earlier with respect to the one elicited by allophonic [ɛ]-[e] contrast, authors suggested that phonemic contrasts are still easier to perceive thanks to a *phonemic* mode of perception which is faster than the *phonetic* mode of perception, which should be employed to perceive allophones belonging to the same phonological category.

Interestingly, allophonic familiarity seems to revert this impairment. Bühler et al. (2017) tested the effects of familiarity with specific allophones in native speakers of Standard German and Swiss German. In a behavioural experiment, the authors measured the discrimination of the allophonic contrasts [t–th] (familiar for Standard German) and [t–tː] (familiar for Swiss German) embedded in a pseudoword by means of a same-different task. Results showed higher discrimination accuracy in Swiss German speakers for the [t–tː] contrast with respect to Standard German ones. While Swiss German listeners could better discriminate the familiar [t–tː] contrast, Standard German listeners' performance was compatible with an assimilated representation of \t\ and \tː\. The [t–th] contrast, instead, appeared to be easily discriminable by both linguistic groups, possibly because of larger acoustic differences. Additionally, in a MMN experiment where the same contrasts embedded in pseudowords were presented in oddball blocks, each group showed smaller MMN effects when the deviant pseudoword contained a familiar allophone suggesting that allophonic familiarity allows listeners to process allophonic contrasts more efficiently.

While contrasting results have been found regarding the discrimination of allophones in isolation (Boomershine et al., 2008; Miglietta et al., 2013; Pallier et al., 1997), in the light of the three theoretical accounts of spoken word recognition taken into consideration, the inference-based account can only explain the allophonic discrimination impairment when allophones are presented in a phonological context as in Peperkamp et al., (2003), as it predicts that variation (i.e., allophonic productions) can be accommodated on the basis of rules depending on the phonological context but cannot account for the discrimination improvement induced by exposure. Instead, the episodic model would predict that phonetic details are always accessible to listeners, in evident contrast with studies showing impairment of allophonic discrimination (Boomershine et al., 2008; Pallier et al., 1997; Peperkamp et al., 2003). Further, the hybrid account predicts that while inferential processes are normally sufficient to accommodate phonetic variation into abstract word representations, listeners would get additional benefits by being exposed to infrequent variants, enhancing lexical access. This model would be appropriate to explain the post-exposure improved allophonic discrimination accuracy of Peperkamp et al. (2003) and the allophonic familiarity effect in Bühler et al. (2017).

Despite allophonic familiarity being beneficial for pseudoword processing (Bühler et al., 2017), it is still unknown whether this profitable relationship could also hold for word processing. According to the inference-based account, allophonic familiarity should exclusively impact on the cost of mapping sounds to abstract pre-lexical units, with no influence on word retrieval per se. Instead, the episodic account would predict that allophonic familiarity could facilitate lexical access, as familiar allophones would be embedded in frequently encountered word episodes stored in memory. Similar predictions can be made for the hybrid account, as allophonic familiarity could still facilitate lexical access to abstract word representations as a function of the frequency of occurrence of the specific word variant.

To study the extent to which listeners are able to process familiar allophonic variations and how this process can mediate the retrieval processes of word representations in different varieties of the same language, we took advantage from patterns of variation displayed in Italian regional varieties (Krämer, 2009). Specifically, linguistic variation is encountered not only across the high number of Italian local dialects, but also in the use that speakers make of the standard language, which in turn is often influenced by the local dialects. Therefore, speakers of Italian may be exposed to up to three varieties: a local dialect, a regional variety of the standard language (regional varieties, for short), and the normative standard language as it is presented, e.g., in the media.

We focused on a specific phonological phenomenon, that is the voicing of sibilant consonants in the Trentino regional variety of Italian. Sibilants do not contrast for voicing in all phonological contexts, nor in all regional varieties of Italian (Krämer, 2003). It has been observed (Bertinetto & Loporcaro 2015) that Central and Southern regional varieties, similarly to the normative Standard conveyed by the media, implement a contrast in terms of voicing word-medially in intervocalic contexts (*fu[s]o*, 'spindle' vs. *fu[z]o* 'melted'), while in Northern regional varieties this contrast is neutralized in favor of the voiced sibilant (*fu[z]o*, 'spindle, melted'). In word-initial prevocalic contexts, sibilants are produced as voiceless in all regional varieties (*sale*, 'salt'). Since pre-consonantally sibilants are assimilated in voicing in all varieties and word-final sibilants occur only in loan-words, this means that voiceless and voiced sibilants have phonemic status in the sound system of Central and Southern regional Standard varieties – albeit only in intervocalic contexts. In Northern regional varieties, voiceless and voiced consonants are in complementary distribution in all contexts.

Of particular interest to our work is the word-medial context where the sibilant is preceded by a sonorant. In this context, regional varieties of Standard Italian generally produce a voiceless sibilant (*sen[s]o*, 'sense'). However, the Trentino variety of Standard Italian – a variety spoken in Trentino, a north-eastern Italian province populated by ~543.000 inhabitants (*Resident Population on 1st January : Provincia Autonoma Trento*, n.d.) –, shows a unique slight deviation from the general production pattern of other Northern varieties of regional Italian. The sibilant following a sonorant consonant is often realized as voiced (e.g., *sen[z]o*, 'sense'), a characteristic also consistent with data collected for dialect surveys. Thus, in the dialect elicitation project VinKo (Rabanus et al., 2021), of 81 participants self-identifying as speakers of a Trentino dialect, 64% (N = 52) pronounced the word *senso* ('sense') with a voiced sibilant as *sen[z]o*, while the remaining 36% (N = 29) pronounced it as *sen[s]o* ($X^2(1, N = 81) = 5.98$, $p < 0.05$). This indicates that this feature of the local dialect has been preserved to some extent also in the regional version of the standard language. We capitalized on this critical difference between Trentino and all other regional Italian varieties and recorded the MMN associated to the presentation of a single word (*senso*, 'sense') embedding either the consonant cluster with the voiced sibilant [nz] typical for the Trentino regional variety, or the voiceless sibilant [ns] that belongs to other productions of regional Italian. In both varieties, the two forms are allophonic in postsonorant position in the sense that no phonemic contrast between voiced and voiceless sibilants is implemented in this context. In general, also in Central and Southern varieties a phonemic contrast between the two sounds arises only in intervocalic contexts, as discussed above. We gathered two groups of Italian native speakers. The first group was formed by participants that were born and always lived in Trentino, to which the voiced sibilant is familiar in this context, and the second group was composed of participants born and raised in the Central and Southern regions of Italy, familiar with the voiceless sibilant.

Starting from the MMN modulations induced by the lexical status of stimuli (Endrass et al., 2004; Pulvermüller et al., 2001, 2004; Shtyrov & Pulvermüller, 2002; Tavano et al., 2012) we generated different sets of predictions for the three different theoretical frameworks. Considering the inference-based account, no variant-specific modulations of the MMN should emerge: the automatic retrieval processes always probe the same word representation despite phonetic variation and allophonic familiarity. In other words, Trentino speakers and Central-Southern speakers should show MMN with comparable amplitude for words embedding the voiced (e.g., *sen[z]o*, 'sense') or the voiceless sibilant (e.g., *sen[s]o*, 'sense'). Episodic models, instead, would predict larger MMN amplitude when listeners hear words embedding the respective familiar allophones signaling the presence of variant-specific word representations. Thus, Trentino speakers should show a larger MMN for words embedding the voiced sibilant (*sen[z]o*), while Central-Southern speakers should show the opposite pattern. Finally, the hybrid approach would predict that words embedding familiar allophones would have an ad-hoc representation in the brain. According to this account, Central-Southern speakers should show no difference in the MMN amplitude between the variant with the voiced sibilant (*sen[z]o*) and that with the voiceless sibilant (*sen[s]o*). Since they have never been exposed to the voiced sibilant in this specific phonological context, they putatively have only one assimilated representation for both variants of the word that is accessed through inferential processes. Differently, Trentino speakers should show a larger MMN for the word embedding the voiced sibilant (*sen[z]o*), as they are familiar with this variant. It is important to note that Trentino speakers are likely to be acquainted with the standard pronunciation (*sen[s]o*) by hearing it in formal and/or educational contexts. Additionally, if familiarity with specific allophonic productions modulate the electrophysiological correlates of memory retrieval (i.e., the enhanced MMN), we expect a relation between the magnitude of the MMN and the degree of exposure to the familiar allophonic production. To further test this hypothesis, self-reported frequencies of production and listening to the Trentino dialect, which is also characterized by the allophonic variation in exam, were collected in Trentino speakers.

2. Methods

2.1 Participants

Eighty-nine healthy right-handed Italian native speakers were recruited (F = 51, M_{age} = 20.96, SD = 3.40). Participants reported to have normal hearing, to be neurologically healthy and not to be under medication that could alter cognitive functioning. Two experimental groups were formed. The Trentino Group (n = 38, F = 21, M_{age} = 22.47, SD = 4.35) was composed of participants born and raised in Trentino, a north-eastern region of Italy, while The Central-Southern Group (n = 51, F = 30, M_{age} = 19.90, SD = 1.97) was composed of participants born in Central-Southern Italian regions (in particular, in the area below the La Spezia-Rimini line[1]) that moved to Trentino not later than 1 month before being tested. Before moving, these participants were never immersed in a Northern Italian speaking environment. All participants were tested in the EEG laboratory of the University of Trento in the Trentino Region. The Central-Southern Group included more participants as they were also involved in a parallel experimental study. Participants in both groups reported to have lived in the same region for at least 10 consecutive years and to have at least 1 parent born in their region of living. Some participants were excluded because of technical problems with the EEG recording devices and the presence of excessive noise in the data (see section *2.4 EEG recording and preprocessing*). Moreover, in order to meet the constraints of the selected framework for

---

[1] According to Maiden & Parry, (1997), the La Spezia-Rimini line consists in an important bundle of isoglosses which divide Western from Eastern Romance languages as well as Northern from Central and Southern Italian dialects.

statistical analyses (see section *2.5 Statistical Analyses*), the Central-Southern group was subsampled to make it identical in numerosity to the Trentino group. The final sample was composed of 30 participants in the Trentino Group (F = 18, Mage = 22.93, SD = 4.62) and 30 participants in the Central-Southern Group (F = 18, Mage = 20.2, SD = 2.24). All participants expressed their informed consent and received (according to their preference) either monetary reimbursement (15 € per session) or university credits for their participation. The study was conducted in line with the Declaration of Helsinki and was approved by the Ethical Committee of The University of Trento (protocol id:2017-26).

2.2 Stimuli

A female Italian native speaker, born, raised, and living in Trentino, was recruited to record the stimuli. The speaker reported to be aware of the peculiar Trentino speakers' production of sibilants with respect to other Italian speakers. The speaker was asked to read the sentence 'Questa cosa non ha senso' (lit. *This thing has no sense,* “This thing makes no sense”) once producing the word 'senso' with the voiceless sibilant [s] after the nasal (Standard Italian), and once with the voiced sibilant [z] (Trentino variant). The target word was placed in broad focus at the end of the sentence, in order to elicit a clearly accented production. The speaker recorded every sentence 3 times. Sentences were recorded at 44100 Hz in a silent room with a professional recorder.

The 3 tokens of each target stimulus (*sen[s]o*, *sen[z]o*) were extracted from the sentences. The tokens were annotated for single phonemes using the software Praat (Paul Boersma & David Weenink, 2018). The duration of each phoneme was measured for all 6 tokens. The voiced and voiceless tokens differed in duration with respect to the word-medial sibilant, with the voiced tokens showing a longer nasal /n/ and a shorter post-nasal /z/ than voiceless tokens. Tokens were re-synthesized using the PSOLA overlap-add algorithm (Moulines & Charpentier, 1990), and the duration of each phoneme was set to the average duration values, calculated for each phoneme across all tokens (Table 1).

Table 1. *Duration of phonemes in milliseconds for each of the initial tokens and average duration*

	Sen[s]o			Sen[z]o			
Phoneme	Token 1	Token 2	Token 3	Token 1	Token 2	Token 3	Average
s	116.8 ms	115.2 ms	107.5 ms	111.9 ms	107.2 ms	108.0 ms	111.1 ms
e	131.3 ms	144.8 ms	136.1 ms	137.6 ms	158.3 ms	134.3 ms	140.4 ms
n	71.3 ms	58.6 ms	79.5 ms	93.6 ms	90.0 ms	100.4 ms	82.2 ms
s - z	133.9 ms	146.7 ms	126.6 ms	76.4 ms	88.5 ms	77.6 ms	108.3 ms
o	177.8 ms	181.5 ms	184.7 ms	188.0 ms	214.5 ms	226.2 ms	195.4 ms

The intensity of all tokens was equalized to an average value of 72 dB. Since there was no stop or silence between the first vowel and the critical phoneme (/s/-/z/), stimuli were not cross-spliced to avoid unnatural transitions in the cross-splicing point (Steinberg et al., 2012). Pitch and intensity profiles along with F1, F2, F3 formants, were inspected to identify possible systematic differences before the onset of the critical phoneme. As expected, all stimuli were highly similar before the onset of /n/. Only tokens of voiced sibilants showed the presence of a pitch contour. The nasal consonant showed a lower pitch frequency in the voiceless than in the voiced tokens. F0 lowering through larynx lowering is an automatic side effect of the articulation of voiced obstruents and is often deliberately extended by speakers to preceding vowels (or, in this case, sonorants), possibly to favor perception of the obstruent as voiced (Kingston, 2011). Spectrograms of the experimental stimuli are available in Supplementary Materials.

2.3 Procedure

Participants were tested individually in a dimly lit room. They were initially asked to fill in a brief questionnaire to collect demographic information (age, gender, educational attainment, geographic origin, place of residence) and language background and to make sure they satisfied the inclusion criteria of the study. After installation of the EEG cap, they were seated in front of a laptop computer and were asked to watch a silent video of our choice while paying no attention to the sounds they heard. Auditory stimulation was delivered by E-Prime 2 software (Schneider & Zuccoloto, 2007) via two speakers at fixed volume (72 dB) positioned at ~40 cm from the participants' ear line while EEG signal was recorded. Two oddball blocks of auditory stimuli were presented. Each block was composed of standard (i.e., frequently presented) stimuli, which were presented 630 times, and deviant (i.e., infrequently presented) stimuli, which was presented 120 times (probability of occurrence = .16). In one block, the Trentino variant *sen[z]o* (voiced sibilant) was used as deviant stimulus and the standard Italian *sen[s]o* (voiceless sibilant) as standard stimulus, while in the other block it was the opposite. The 3 tokens for *sen[z]o* and the 3 token for *sen[s]o* were equiprobably presented both as standard and as deviant stimuli. In the block in which *sen[s]o* was presented as standard stimulus and *sen[z]o* as deviant, in each standard trial one of the 3 tokens of *sen[s]o* was presented with identical probability across tokens. The same logic was used for deviant *sen[z]o* and likewise for the block in which the standard/deviant status was reversed (i.e., *sen[z]o* standard and *sen[s]o* deviant). Each stimulus lasted 680 ms and was played one after the other, with an interstimulus interval (ISI) of 418 ms. Within each block, standard and deviant stimuli were randomly presented, with the constraint that at least two standards had to occur before each deviant. Each block lasted approximately 15 minutes with a small break between the two. The order of the blocks was counterbalanced across participants. Each experimental session lasted approximately 1 hour per participant: about 30 minutes for preparation and 30 minutes for the experiment.

After the EEG session, the spontaneous production of the sibilant in the participants of the Trentino Group was evaluated, by asking them an apparently unrelated question that could elicit the production of a word containing the consonant cluster /n/+/s/ ("Per salire al terzo piano di un palazzo, puoi prendere le scale oppure…?", "If you need to go to the third floor of a building you can take the stairs or…?"; Answer: "l'ascensore", 'The elevator'). This allowed us to assess whether they spontaneously produced the /s/ phoneme either as voiceless or voiced. Finally, participants of the Trentino Group were asked to fill in a brief Sociolinguistic Questionnaire to investigate the frequency of speaking and listening to the Trentino dialect: They were asked to express on a 1(never) to 5 (always) points Likert scale how frequently they speak or listen to the Trentino dialect with family members and friends – scores for speaking and listening were separately collected. The questionnaire is available in the Supplementary Materials.

2.4 EEG recording and preprocessing

The EEG signal was acquired with an eego sports system (ANT Neuro) at a sampling rate of 1000 Hz, from 64 Ag/AgCl shielded electrodes (ANT Neuro Waveguard Cap) placed on the scalp according to the 10-10 international electrode positioning system and referenced to the CPz site. Electro-oculograms were recorded with an additional EOG electrode placed under the left eye. Impedance of each electrode was kept below 20 kΩ. Four participants of the Trentino group were excluded due to a failure of the EEG recording device. EEG data were re-referenced to average reference (excluding EOG and mastoid channels), resampled at 250 Hz, and digitally filtered with a 50 Hz notch filter and a passband Butterworth filter (0.01-30 Hz, Order 4). Independent Component Analysis with ICA Infomax algorithm (Amari et al., 1996) was run on the continuous signal and components corresponding to eye blinks were visually identified and removed. Epochs were extracted in the -200 ms prior and 800 ms post word onset and baseline correction was performed using the whole pre-stimulus interval. Epochs were time-locked to word onset as acoustic-phonetic differences originating from co-articulation and voicing could shift the onset of the MMN in time from the predicted time point (i.e., at 333 ms, corresponding to the onset of the allophone of interest [s]/[z]). Epochs with signal amplitude exceeding a [-100 μV 100 μV] threshold in any channel were rejected to remove excessively noisy epochs. After this procedure, 4 participants from the Trentino group and 4 participants from the Central-Southern group were excluded from the analysis as they showed less than 100 deviant artifact-free epochs for each condition. The preprocessing procedures were performed using MATLAB toolboxes EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014).

2.5 Statistical Analyses

Differences between groups and conditions were evaluated using a nonparametric cluster-based permutation approach. This approach was preferred to the parametric one because it allowed us to perform greater amounts of statistical comparisons along different electrode sites and time bins while still being sure of controlling the family-wise error rate (FWER). This need was imposed by the nature of the stimuli we deployed: even though speech stimuli were matched as much as possible along different physical dimensions, it is difficult to fully account for small idiosyncrasies in each token, which could affect the spatio-temporal characteristics of the component of interest.

In this approach, data points with a p-value < .05 (critical alpha level, two-tailed) are selected and clustered on the basis of temporal and spatial adjacency. Cluster statistics were calculated by summing all the t-values within any identified cluster. The distribution of t-values under the null hypothesis was computed by calculating the test statistic several times (N = 10,000) on random partitions of the data shuffled across conditions. The proportion of random partitions where the observed t-value is larger than the t-value drawn from the permutation distribution represents the cluster p-value (Maris & Oostenveld, 2007). When independent samples (i.e., groups) are compared with this method, the dimensions of each sample must meet. For this reason, the Central-Southern (n = 47) was randomly subsampled to match the size of the Trentino group after preprocessing (n = 30).

Time-locked ERP responses were calculated within each individual participant for voiceless and voiced stimuli for all conditions: *sen[s]o* standard, *sen[s]o* deviant, *sen[z]o* standard, *sen[z]o* deviant. The MMN was calculated across blocks by subtracting the standard ERPs from the deviant ERPs within the same stimulus type: MMN *sen[s]o* = *sen[s]o* deviant - *sen[s]o* standard; MMN *sen[z]o* = *sen[z]o* deviant - *sen[z]o* standard. The aim of this computation, which is extensively used in the literature (Eulitz & Lahiri, 2004; Fu & Monahan, 2021; Hestvik & Durvasula, 2016; Jacobsen, Schröger, & Alter, 2004; Jacobsen, Schröger, & Sussman, 2004; Peter et al., 2010; Steinberg et al., 2010) is to reduce the effect of physical differences that occur between standard and deviant stimuli and to isolate the effect of the cognitive process of interest. Presenting the same deviant stimulus as standard in another block, allows to record the exogenous activity related to that specific stimulus that can be subtracted out from the deviant ERP.

Cluster-based permutation tests were implemented using the MATLAB toolbox FieldTrip (Oostenveld et al., 2011). The signal amplitude of the ERPs across conditions (*sen[s]o* standard vs *sen[s]o* deviant and *sen[z]o* standard vs *sen[z]o* deviant) was compared by multiple t-tests within each experimental group, performed at each data point in a subset of channels containing Pre-frontal, Frontal, Fronto-Central and Central electrode sites (Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, Cz, C4, F5, F1, F2, F6, FC3, FCz, FC4, C5, C1, C2, C6) where MMN is typically distributed (Näätänen et al., 2007). The amplitude of MMN responses for *sen[s]o* and *sen[z]o* was then compared within groups in the largest time window where deviant ERPs significantly differed from standard ERPs. Interaction effects between groups and word variant were evaluated confronting the difference obtained by subtracting signal amplitude of MMN *sen[s]o* from the one of MMN *sen[z]o* between groups. MMN peaks were identified as the most negative points, in the 200-800 ms time window. Peak latency was measured within each combination of group and word variant by averaging the peak latency value of each individual participant across all channels in the channel pool reported above.

The self-reported frequencies of speaking and listening to dialect were analyzed by means of an Ordinal Logistic Regression Model using the package “MASS” (Venables & Ripley, 2002) in

R Software (R Core Team, 2013). Data were fitted to the full model with fixed factors of activity (speaking, listening), context (friends, family), sex (female, male) and their interactions. The best model was selected via likelihood-ratio Chi-squared tests performed with the drop1 R function. The p-values of the reported effects were calculated by comparing the associated t-value with the normal distribution.

The amplitudes of the MMN for both *sen[s]o* and *sen[z]o* were evaluated for each participant of the Trentino group as the average across the electrode sites and the time bins that formed a significant cluster in the comparison between the two MMN waveforms. The obtained values of the MMN amplitude were tested for correlations with the responses of each item of the Sociolinguistic Questionnaire by calculating the Kendall rank correlation coefficient (Abdi, 2007). P-values of the correlation tests were corrected with the False Discovery Rate correction (Benjamini & Hochberg, 1995).

3. Results

3.1 EEG data

A sustained MMN response after word onset was successfully elicited in every group and for every word variant, mainly distributed across Frontal and Fronto-Central electrode sites. The Trentino group showed a significant negative cluster (p < .001), indicating a negative ERP effect of *sen[z]o* deviant with respect to *sen[z]o* standard in the 392-788 ms time window (peak at 518 ± 153 ms) and of *sen[s]o* deviant with respect of *sen[s]o* standard in the 288-796 ms time window (peak at 507 ± 147 ms). The Central-Southern group also showed a significant negative cluster (p = .002) in the 316-700 ms time window between *sen[z]o* standard and *sen[z]o* deviant (peak at 527 ± 145 ms) and in the 416-684 ms time window between *sen[s]o* standard and *sen[s]o* deviant (p < .001, peak at 502 ± 166 ms). When looking at within-group differences the analysis showed a significant cluster (p = .020) for Trentino speakers only, approximately between 528 and 648 ms and predominately distributed over frontal and frontocentral right channels (F6, F8, FC6): MMN response was larger for voiced (familiar) than for voiceless (less familiar) stimuli on frontal right electrodes[2]. No significant clusters were found in the between-group analyses. Results are summarized in Figure 1.

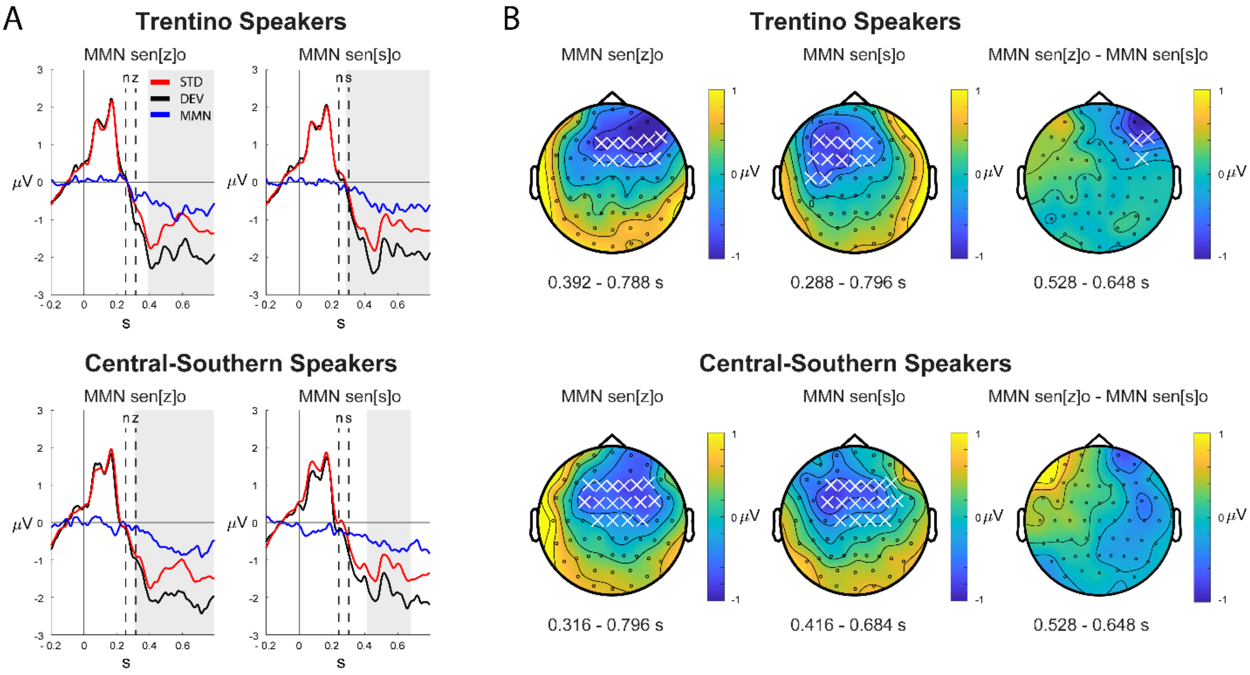

Figure 1. MMN to *sen[s]o* and *sen[z]o* for the Trentino and the Central-Southern speakers' group. (A) MMN (blue) is plotted for the Fz site within every group and for every word variant by subtracting the standard ERP (red) from the deviant ERP (black). The time-window in which significant clusters were found is represented by light grey areas. Vertical dashed lines indicate the onset of /n/ and [s]-[z] respectively. (B) Topographies show the spatial distribution of the MMN in

---

[2] It is worth noting that the topography of the differential effects between the two MMNs depicted in Figure 1 (5th column) might resemble the topography of an independent component of horizontal eye movements, as noted by an anonymous reviewer. Considering that the topographies of MMN sen[z]o and MMN sen[s]o are lateralized towards opposite directions in both groups, if they were contaminated by eye-movements artifacts they would reflect saccadic activity directed towards opposite directions on the basis of the condition. Although some saccades probably occurred during the experiment (since participants were free to move their gaze and visual attention towards different part of the screen while the silent movie was played) it is highly unlikely that they could be linked and time-locked to our auditory stimulation, as they were mostly elicited by visual stimulation. In fact, while the visual stimulation could likely elicit saccades towards random directions at any given time-point, the auditory stimulation was delivered with equal intensity from a left and a right speaker irrespectively of the condition. Therefore, it is safe to assume that auditory stimulation could not systematically elicit saccades time-locked with auditory events directed towards different directions on the basis of the condition.

the time windows indicated below the maps that correspond to the temporal extension of the cluster. Electrodes that were included in the clusters for more than 50% of the samples within the cluster time windows are represented by white marks superimposed to the maps. The topographical map representing the difference between the two MMN waveforms in Central–Southern speakers (2^{nd} row, rightmost plot) refers to the time window in which a significant cluster is found for Trentino speakers only for illustrative purposes.

3.2 Sociolinguistic Questionnaire and Production data

The data of the sociolinguistic questionnaire administered to the Trentino Group are summarized in Table 2. The final Ordinal Logistic Regression Model predicted the rate of dialect use as a function of activity (speaking and listening), context (friends and family) and sex (female, male) as fixed factors. The model showed significant effects of context ($\beta = 1.89$, SE = 0.44, $t = 4.21$, $p < .001$) and activity ($\beta = 0.94$, SE = 0.34, $t = 2.75$, $p = .006$), indicating that participants reported to listen to dialect more likely than to speak it and to listen to or speak dialect more likely with family members than with friends. All the Trentino participants included in the final sample, spontaneously produced the critical sibilant phoneme as voiced.

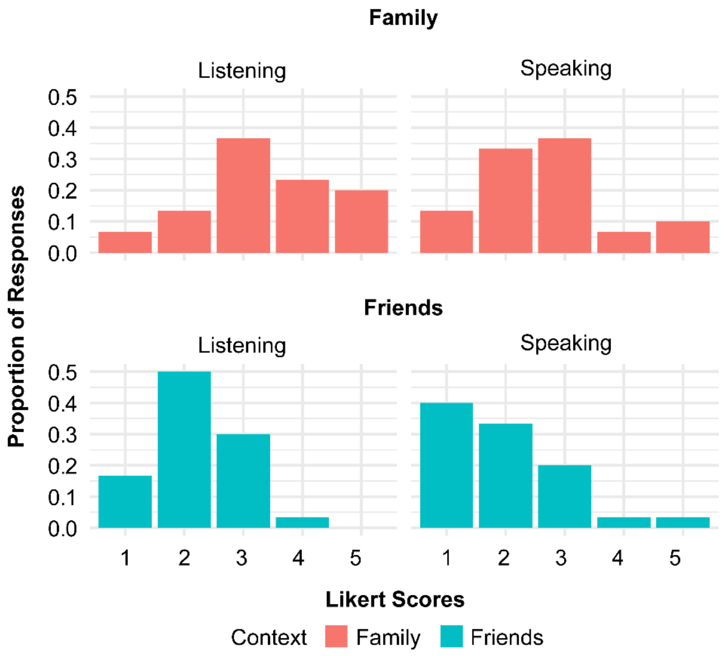

Figure 2. Proportion of Likert Scores of the Sociolinguistic Questionnaire divided by Context (Family, Friends) and Activity (Listening, Speaking).

3.3 Correlations

There was a marginally significant positive correlation between the amplitude of MMN to *sen[s]o* and the self-reported frequency of listening to dialect in family contexts ($\tau = .34$, $p = .054$): The more participants reported to be passively exposed to dialect in family context, the weaker (i.e., less negative) the MMN elicited by the deviant word embedding the Standard Italian voiceless sibilant (Figure 2). No further correlation approached significance (all ps > .24).

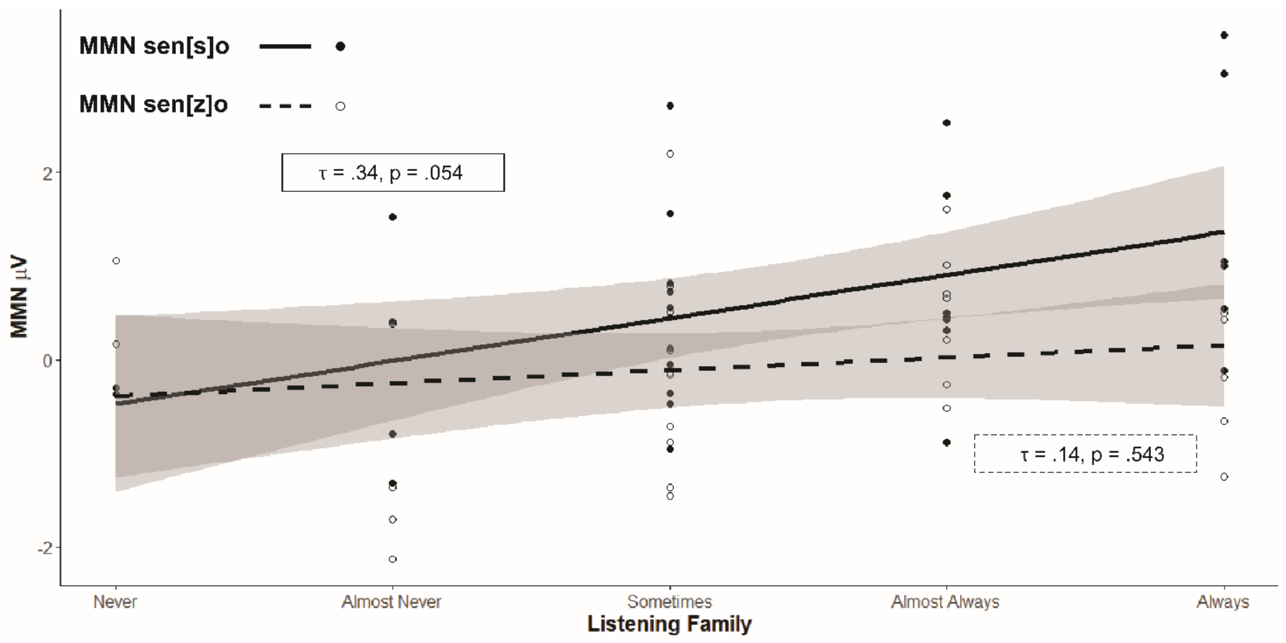

Figure 3. Correlation between the amplitude of MMN and the self-reported frequency of listening to dialect from family members. Black dots represent individual observations of the amplitude of MMN to *sen[s]o* in function of the self-reported frequencies of listening to dialect from family members; black solid line represents the slope of the correlation. White dots represent individual observations of the amplitude of MMN to *sen[z]o* in function of the self-reported frequencies of listening to dialect from family members; black dashed line represents the slope of the correlation. The Gray areas represents C.I.

4. Discussion

This ERP study investigated how allophonic variation and familiarity with specific allophones influence automatic memory retrieval processes for words. Native speakers of the Trentino regional variety of Italian and of Central-Southern Italian varieties took part in the study. We tested the MMN elicited by the presentation of the word *senso* embedding either the voiced sibilant (*sen[z]o*), typical of the Trentino regional Italian, or the voiceless sibilant (*sen[s]o*) of Standard Italian and of the Central-Southern varieties of Italian. Additionally, measures of self-reported frequency of use of the Trentino dialect were collected to investigate the impact of exposure to the allophonic variant on the electrophysiological response of the Trentino speakers.

The ERP data showed that a sustained MMN response was successfully elicited for both deviant *sen[s]o* and *sen[z]o* in both groups, mainly distributed across Frontal and Fronto-Central electrode lines. The deviant ERP showed a clear negative displacement from the standard ERP (on average) from 300 ms until 800 ms after word onset, as indicated by the significant clusters. The peak latency of the MMN is overall consistent with the perception of a phonetic difference in the time window of the critical sibilant (i.e., 150-200 ms after sibilant onset; Näätänen et al., 2007). The successful elicitation of the MMN response indicated that both Trentino and Central-Southern speakers pre-attentively detected the phonetic dissimilarities occurring between the two variants of the word. This may be apparently in contrast with the behavioural results by Peperkamp et al. (2003), who found that allophonic discrimination is more difficult when allophones were embedded in a non-lexical but phonologically legal context. However, electrophysiological measures are more sensitive to the processing of small acoustic differences that may not be detected by behavioural measures (van Zuijen et al., 2006). Our finding thus suggests that listeners pre-attentively perceive allophonic variations embedded in meaningful words.

Within group comparisons revealed that, in the Trentino speakers' group, the familiar *sen[z]o* elicited a larger MMN response at the rightmost Frontal and Fronto-central electrode sites than the unfamiliar *sen[s]o*, while no amplitude differences between the MMN elicited by *sen[s]o* and *sen[z]o* were found for the Central-Southern speakers' group. However, between group comparisons indicated that the amplitude difference between MMN *sen[s]o* and MMN *sen[z]o* found for Trentino speakers was not statistically different from the one computed for Central-Southern speakers. This result may suggest that both groups do not have differentiated word representations for *sen[s]o* and *sen[z]o*, despite the latter supposedly being the most frequent production for Trentino speakers and the actual version spontaneously produced by the Trentino participants of the present study. The inference-based account for word recognition (Gaskell & Marslen-Wilson, 1998) would predict that listeners can accommodate phonetic variability – hence allophonic variations – via inferential processes that follow the rules dictated by the phonological context. We specifically predicted that if this model was the best one to explain the perception of allophonic variation in word contexts, no variant-specific amplitude modulation of the MMN would have emerged. To this regard, previous studies showed that the MMN is larger when the deviant stimulus belongs to the listener's native language (Dehaene-Lambertz, 1997) and form a meaningful word (Shtyrov & Pulvermüller, 2002). Authors interpreted this effect as an index of an automatic memory trace retrieval process for native phonemes or known words, but the presence of this effect does not emerge from our between-group analyses. While both groups have shown to perceive the [s]-[z] contrast in a word context (as indicated by the presence of a clear MMN for both word variants), participants might not have differentiated word representations for *sen[s]o* and *sen[z]o* and simply access the same abstract *senso* representation by accommodating allophonic variation thanks to inferential processes.

The absence of between-group effect is critical for the interpretation of results. However, considering the patterns at the within-group level – which showed a significant difference between MMN *sen[s]o* and MMN *sen[z]o* for Trentino speakers but not for the Central-Southern speakers – one possibility is that, although our samples were relatively large, our study lacked the sufficient power to highlight a between-group difference given a) the extremely conservative nature of the statistical approach we employed and b) the intrinsic weakness in term of statistical power of the between group analysis – as a matter of fact, any linear combination of ERP data has the advantage of removing non relevant aspects of the EEG, but at the same time strongly affects signal-to-noise ratio (Luck, 2014). With respect to this last consideration, it must be noted that the between-group comparison is relative to a difference (group) in a difference (MMN *sen[s]o* and MMN *sen[z]o*) of a differential effect (deviant minus standard). Therefore, with this possibility in mind, we now attempt to tentatively discuss the results of the within-group analyses. Looking at the results for Trentino speakers only, it might be possible that larger amplitude MMN may be elicited also when words embed familiar allophones with respect to when they do not. Native speakers of the Trentino variety are exposed to both allophonic variants as they can hear Standard Italian *sen[s]o* in educational and institutional contexts or through media, but they can putatively hear Trentino *sen[z]o* more frequently, especially in family contexts or with friends. Being frequently exposed to the word variant *sen[z]o*, native speakers of the Trentino variety may have built differentiated memory traces for the familiar *sen[z]o* and the less familiar *sen[s]o*. Conversely, Central-Southern speakers were never exposed to the Trentino variety prior to participating in the experiment. Thus, it is unlikely that they developed two separated word representations for *sen[s]o* and *sen[z]o*, while still being able to discriminate the phonetic differences between the two.

One similar study investigated whether native speakers of Standard American English (SAE) could pick up dialect specific phonetic features by comparing the MMN elicited by the word *hello* produced in SAE dialect and the MMN elicited by the same word produced in African American English dialect (AAE) (Scharinger et al., 2011). Results showed a reliable elicitation of the MMN for both versions of the word while no MMN was found in a control condition in which the standard-deviant acoustic differences were acoustically matched with the ones characterizing the condition in which SAE and the AAE stimuli were presented. Authors argued that while acoustic differences may have had a role in determining the elicitation of the MMN to SAE and AAE stimuli, since no MMN emerged in the control condition, results mainly reflect a top-down modulation induced by dialectal knowledge in long-term memory. Moreover, authors showed that the MMN to SAE stimuli (which were familiar to SAE participants) was larger than that elicited by unfamiliar AAE stimuli. This specific result is in line with the withing-group results of the present study for Trentino speakers, showing a larger amplitude MMN for the familiar *sen[z]o* vs the unfamiliar *sen[s]o*.

With respect to the three theoretical accounts of word recognition, always bearing in mind the absence of between-groups differences, the patterns emerging at the within-groups level of Trentino and Central-Southern speakers might be compatible with the hybrid account of word recognition. This class of models predicts that specific abstract representations for frequent word variants may be developed to reduce the impact of word retrieval on cognitive resources (Connine & Pinnow, 2006; Pinnow & Connine, 2014) as a function of the quantity of exposure with the specific word variants (Sumner & Samuel, 2009). That is, the access to word variants representations is weighted by the frequency of occurrence: Trentino speakers might have two separated word representations for *sen[s]o* and *sen[z]o*, with the representation for *sen[z]o* having stronger activation weights, given that this word variant is encountered more frequently with respect to *sen[s]o*. Hence, they showed a larger amplitude MMN to the more frequent *sen[z]o*, a result in line with another study with Russian speakers in which MMN to frequent vs infrequent words showed larger amplitude (Aleksandrov et al., 2017). Instead, Central-Southern speakers did not show the opposite pattern even if they are more familiar with *sen[s]o* (Standard Italian). This pattern can also be explained through the lens of the hybrid approach, which does not exclude the usage of inferential processes to accommodate allophonic variation. When encountering the unfamiliar allophone [z] in a context where [s] is expected, Central-Southern speakers might have assimilated the [z] allophone within the native phonological /s/ category, thus retrieving the “standard form” (i.e., an abstract representation) of the word senso.

In contrast with the interpretation of the between-group analysis, the pattern of within-group results would not fit completely with the inference-based account. This would only be apt to explain the results for Central-Southern speakers which may have accommodated the allophonic variation into a single abstract representation of the word *senso* (‘sense’) but not the results for Trentino speakers. The episodic account could still explain the results for Trentino speakers, as it predicts the existence of frequency-weighted episodic memories for all the encountered word variants, but it would not be apt to explain the results for Central-Southern speakers as if they only have experienced the *sen[s]o* variant, they would have shown a stronger MMN for that specific word. The interpretation of the within-group results is also in line with Sebastián-Gallés, Vera-Constán, Larsson, Costa, & Deco (2009) who show that Catalan listeners form differentiated lexical representations for words spoken in a Spanish dialect as a result of prolonged exposure, while phonemic categories are not affected. Catalan listeners recognize /e/ and /ɛ/ as separate phonemes, while Spanish listeners assimilate them in /e/. In this study, Catalan listeners showed a N400 for words vs non-words contrasting only for /e/ - /ɛ/ vowels but did not show any effect when the contrast was /ɛ/ - /e/ and the non-word containing /e/ was a recognized word variant in Spanish. In addition, the amplitude of MMN for /de/ - /dɛ/ contrast in isolation did not reveal any difference, suggesting no violation of phonemic boundary.

It is important to note that the right topographical distribution of the difference between the two MMNs for Trentino Speakers slightly diverges from the typical distribution of the enhancement effect, which is usually more evident on the midline electrodes (Pulvermüller & Shtyrov, 2006). The reason for this topographical inconsistency may lie in the involvement of phonetic analysis processes. Bühler, Schmid, and Maurer (2017) showed that when familiar vs unfamiliar allophones are embedded in pseudowords, the MMN is weaker for familiar allophones. Moreover, source reconstruction suggests the right hemisphere as a possible source of the effect. Authors suggest that the activity of the right-lateralized sources is linked to a stronger need of non-linguistic phonetic analysis of unfamiliar sounds that can impact linguistic processes. This topographical distribution of the effect converges with the one reported in the scalp topographies of our study, yet the direction of the effect at the ERP level seems to differ. The origin of this divergence may lie in the different role the right-lateralized processes would undertake when meaningful words are presented to the listener. In fact, during word perception, the right hemisphere is involved in acoustic/phonetic analysis that supports left-hemispheric phonemic processes and its involvement seems to be facilitated by lexical context (Wolmetz et al., 2011). Bühler, Schmid et al. (2017) suggest that when pseudowords are presented in the experiment, right-lateralized processes reflect non-linguistic phonetic analysis that is still facilitated by allophonic familiarity. However, when meaningful words are presented, lexical and phonetic information could be mutually beneficial. In this way, while the output of right-lateralised phonetic processes can be channelled into a word form, the word recognition system can finally encode familiar phonetic information and retrieve the appropriate representation. Following the hybrid account of word recognition models, these results could indicate that specific phonetic representations are formed for frequent allophonic productions and their activation may also rely on right-lateralized processes.

An additional clarification about the role of exposure to specific word variants comes from the results of the correlation analyses. The marginally significant correlation between the self-reported frequency of listening to the Trentino dialect in familiar context and the MMN *sen[s]o* suggests that higher frequency of exposure to the dialect in Trentino speakers was associated to smaller MMN to the unfamiliar *sen[s]o* on frontal and frontocentral right electrodes, while no correlation was found for the MMN to the familiar *sen[z]o*. Moreover, as indicated by the regression analyses, both passive and active exposure to dialect were more likely to happen in familiar context. This suggests that the more individual listeners are exposed to a specific word embedding familiar allophones, the lesser the phonetic-related processes are involved when standard phonology is heard.

4.1 Final remarks and conclusions.

A critical aspect of this study clearly relates to the inconsistencies between the interpretation of the results based on the between-group effect and the one emerging from the within-group effects. While the between-group results could be framed by the inference-based accounts of word recognition, the within-group results as well as the correlation analyses might suggest that the hybrid approach would be more suited. To this regard, it must be acknowledged that the link between our data and spoken word recognition models is indirect as it is basically grounded on the hypothesis that MMN electrophysiological response is strongly dependent on long term memory representation of spoken words. This link is supported by extensive empirical data on this topic (Endrass et al., 2004; Pulvermüller et al., 2001, 2004; Shtyrov & Pulvermüller, 2002; Tavano et al., 2012), showing a larger amplitude MMN for word stimuli with respect to phonologically balanced non-words. However, the difference between *sen[s]o* and *sen[z]o* is allophonic, thus possibly more fine-grained with respect to a definite lexical contrast between words and non-words. In fact, a crossover interaction could have been predicted only if lexical representation of spoken words would be firmly linked to the way a word is produced. Therefore, this caveat possibly makes our paradigm rather suboptimal for strong inferential conclusions about models of spoken word recognition which revolve around lexical access. Further studies on this matter should possibly use very large sample sizes as well as try to include additional behavioural measures (e.g., discrimination and categorization tasks) and sociolinguistic questionnaire that could guide the interpretation of the electrophysiological measures both at the between- and within-group level.

Another potential limit of our study is the occurrence of an early onset of the MMN response, which suggests that words could be pre-attentively discriminated slightly before the onset of the /s/ phoneme. The F0 lowering on the nasal consonant preceding the onset of the voiced sibilant may have served as additional phonetic cue to signal the upcoming allophonic variation. This particular cue which is generated by an automatic process may further be considered as a proper part of the whole allophonic variation. However, the peak latencies of the MMN response suggest that the presentation of the critical sibilant phoneme still generated the strongest negative peak amplitudes in the deviant ERPs. When multiple phonetic violations of the standard word representation in short-term memory occur, MMN can also appear with contingent multiple peaks (Truckenbrodt et al., 2014). While we have not been able to statistically address the detection of multiple MMN peaks at single subject level, the grand-average plots suggest the possibility that our average MMN are built up by multiple peaks that may be due to the detection of multiple deviant features in each single token in slightly different time points. To avoid possible confounds and misinterpretations, the presence of statistically significant differences located outside the temporal bounds of the MMN peak consistent with the presentation of the /s/-/z/ phoneme were treated with caution.

Additionally, it is worth acknowledging that while the current implementation of the MMN paradigm nicely accounts for the influence of physical features of the stimuli on the MMN waveform, it cannot avoid alleged contaminations due to the different direction of change detection between standard and deviant stimuli. In fact, while MMNs were computed by subtracting the ERPs of each standard stimulus by the one elicited by the same stimulus presented as deviant, both deviants occurred in different “standard contexts” (i.e., the deviant sen[z]o after standard sen[s]o, and the deviant sen[s]o after standard sen[z]o). Future studies might develop new implementations of the MMN paradigm that could both control for the effects due to physical features of the single stimuli and for the ones stemming from possible differences related to the direction of change between the stimuli.

In conclusion, by capitalizing on multilingual experience of people speaking Italian and different regional varieties, we suggest that words embedding familiar allophones and words embedding standard phonemes are differently represented in the brain of native speakers of a specific regional variety. At the electrophysiological level, this difference may be characterized by the additional involvement of specific right-lateralized processes of phonetic analysis that enrich word representations with familiar phonetic information, supporting the hybrid account of word recognition. Moreover, the strength of activation of such processes seems to be modulated by the individual degree of exposure with allophonic word forms.

Acknowledgements: The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465, “AThEME”.

Declaration of Competing Interest: none.

Data Statement: The datasets acquired in the context of the present study are not publicly available because consent for publication was not obtained from the participants. Data and scripts are available from the corresponding author on reasonable request.

References

Abdi, H. (2007). The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, 508–510.

Amari, S., Cichocki, A., & Yang, H. H. (1996). *A new learning algorithm for blind signal separation*. 757–763.

Avery, P., Dresher, B. E., & Rice, K. (2008). *Contrast in phonology: Theory, perception, acquisition* (Vol. 13). Walter de Gruyter.

Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(02), 131. https://doi.org/10.1017/S0025100305002148

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. *Contrast in Phonology: Theory, Perception, Acquisition*, *13*, 145–172.

Bühler, J. C., Schmid, S., & Maurer, U. (2017). Influence of dialect use on speech perception: A mismatch negativity study. *Language, Cognition and Neuroscience*, *32*(6), 757–775. https://doi.org/10.1080/23273798.2016.1272704

Calabrese, A. (2012). Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception. *Journal of Neurolinguistics*, *25*(5), 355–381. https://doi.org/10.1016/j.jneuroling.2011.03.005

Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (2002). *The handbook of language variation and change*. Wiley Online Library.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, *23*(3), 235–245.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*(4), 919–924.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Eisner, F., & McQueen, J. M. (2018). Speech perception. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *3*, 1–46.

Endrass, T., Mohr, B., & Pulvermuller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, *19*(6), 1653–1660. https://doi.org/10.1111/j.1460-9568.2004.03247.x

Eulitz, C., & Lahiri, A. (2004). Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition. Journal of Cognitive Neuroscience, 16(4), 577–583. https://doi.org/10.1162/089892904323057308

Fasold, R. W., & Connor-Linton, J. (2014). *An introduction to language and linguistics*. Cambridge university press.

Fu, Z., & Monahan, P. J. (2021). Extracting Phonetic Features From Natural Classes: A Mismatch Negativity Study of Mandarin Chinese Retroflex Consonants. *Frontiers in Human Neuroscience*, *15*, 609898. https://doi.org/10.3389/fnhum.2021.609898

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380. https://doi.org/10.1037/0096-1523.24.2.380

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251. https://doi.org/10.1037/0033-295X.105.2.251

Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, *152*, 28–43.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, *41*(4), 654–659. https://doi.org/10.1111/1469-8986.2004.00175.x

Jacobsen, T., Schröger, E., & Sussman, E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research*, *20*(3), 473–479. https://doi.org/10.1016/j.cogbrainres.2004.03.021

Kingston, J. (2011). Tonogenesis: Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–30). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444335262.wbctp0097

Krämer, M. (2003). Variation of s-voicing in two varieties of Italian. *Proceedings of the XXVIIIth Incontro Di Grammatica Generativa. Lecce: Congedo Editore*.

Krämer, M. (2009). *The phonology of Italian*. Oxford University Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431. https://doi.org/10.1037/h0020279

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Maiden, M., & Parry, M. (1997). Introduction. In *The dialects of Italy* (p. 3). Routledge.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Miglietta, S., Grimaldi, M., & Calabrese, A. (2013). Conditioned allophony in speech perception: An ERP study. *Brain and Language*, *126*(3), 285–290. https://doi.org/10.1016/j.bandl.2013.06.001

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361. https://doi.org/10.1016/j.cognition.2013.07.011

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5–6), 453–467. https://doi.org/10.1016/0167-6393(90)90021-Z

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6–18.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*, *42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432–434. https://doi.org/10.1038/385432a0

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: A critical review and reinterpretation. *Biological Psychology*, *8*(2), 81–136. https://doi.org/10.1016/0301-0511(79)90053-X

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. https://doi.org/10.1016/S0010-0277(97)00030-9

Paul Boersma & David Weenink. (2018). *Praat: Doing phonetics by computer [Computer program]* (Version 6.0.37) [Computer software]. http://www.praat.org/

Peperkamp, S., Pettinato, M., & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. *Proceedings of the 27th Annual Boston University Conference on Language Development*, *2*, 650–661.

Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, *482*(1), 71–75.

Pierrehumbert, J. B. (2006). The statistical basis of an unnatural alternation. *Laboratory Phonology*, *8*, 81–107.

Pinnow, E., & Connine, C. M. (2014). Phonological variant recognition: Representations and rules. *Language and Speech*, *57*(1), 42–67. https://doi.org/10.1177/0023830913479105

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory Traces for Words as Revealed by the Mismatch Negativity. *NeuroImage*, *14*(3), 607–616. https://doi.org/10.1006/nimg.2001.0864

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*(1), 49–71. https://doi.org/10.1016/j.pneurobio.2006.04.004

Pulvermüller, F., Shtyrov, Y., Kujala, T., & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, *41*(1), 106–112. https://doi.org/10.1111/j.1469-8986.2003.00135.x

R Core Team. (2013). *R: A language and environment for statistical computing*.

Resident population on 1st January: Provincia Autonoma Trento. (n.d.). Retrieved November 20, 2020, from http://dati.istat.it/Index.aspx?lang=en&SubSessionId=86772a9d-95e2-4b63-98aa-00f55af58157

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2011). You had me at “Hello”: Rapid extraction of dialect information from spoken words. *NeuroImage*, *56*(4), 2329–2338. https://doi.org/10.1016/j.neuroimage.2011.04.007

Schneider, E., & Zuccoloto, A. (2007). E-prime 2.0 [Computer software]. *Pittsburg, PA: Psychological Software Tools*.

Sebastián-Gallés, N., Vera-Constán, F., Larsson, J. P., Costa, A., & Deco, G. (2009). Lexical Plasticity in Early Bilinguals Does Not Alter Phoneme Categories: II. Experimental Evidence. *Journal of Cognitive Neuroscience*, *21*(12), 2343–2357. https://doi.org/10.1162/jocn.2008.21152

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport*, *13*(4), 521–525.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2010). Preattentive Phonotactic Processing as Indexed by the Mismatch Negativity. *Journal of Cognitive Neuroscience*, *22*(10), 2174–2185. https://doi.org/10.1162/jocn.2009.21408

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study. Misleading formant transitions hinder automatic phonological processing. *The Journal of the Acoustical Society of America*, *131*(4), 3120–3140. https://doi.org/10.1121/1.3688515

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. https://doi.org/10.1016/j.jml.2009.01.001

Tavano, A., Grimm, S., Costa-Faidella, J., Slabu, L., Schröger, E., & Escera, C. (2012). Spectrotemporal processing drives fast access to memory traces for spoken words. NeuroImage, 60(4), 2300–2308. https://doi.org/10.1016/j.neuroimage.2012.02.041

Trubetzkoy, N. S. (1969). *Principles of phonology.*

Truckenbrodt, H., Steinberg, J., Jacobsen, T. K., & Jacobsen, T. (2014). Evidence for the role of German final devoicing in pre-attentive speech processing: A mismatch negativity study. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01317

van Bezooijen, R., & Gooskens, C. (1999). Identification of Language Varieties: The Contribution of Different Linguistic Levels. *Journal of Language and Social Psychology*, *18*(1), 31–48. https://doi.org/10.1177/0261927X99018001003

van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Näätänen, R., & Tervaniemi, M. (2006). Implicit, Intuitive, and Explicit Knowledge of Abstract Regularities in a Sound Sequence: An Event-related Brain Potential Study. *Journal of Cognitive Neuroscience*, *18*(8), 1292–1303. https://doi.org/10.1162/jocn.2006.18.8.1292

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (4th ed.). Springer. http://www.stats.ox.ac.uk/pub/MASS4/

VinKo ('Varieties in Contact'). (n.d.). Retrieved January 10, 2021, from https://www.vinko.it/index.php?lang=en

Wolmetz, M., Poeppel, D., & Rapp, B. (2011). What Does the Right Hemisphere Know about Phoneme Categories? *Journal of Cognitive Neuroscience*, *23*(3), 552–569. https://doi.org/10.1162/jocn.2010.21495

Supplementary Materials

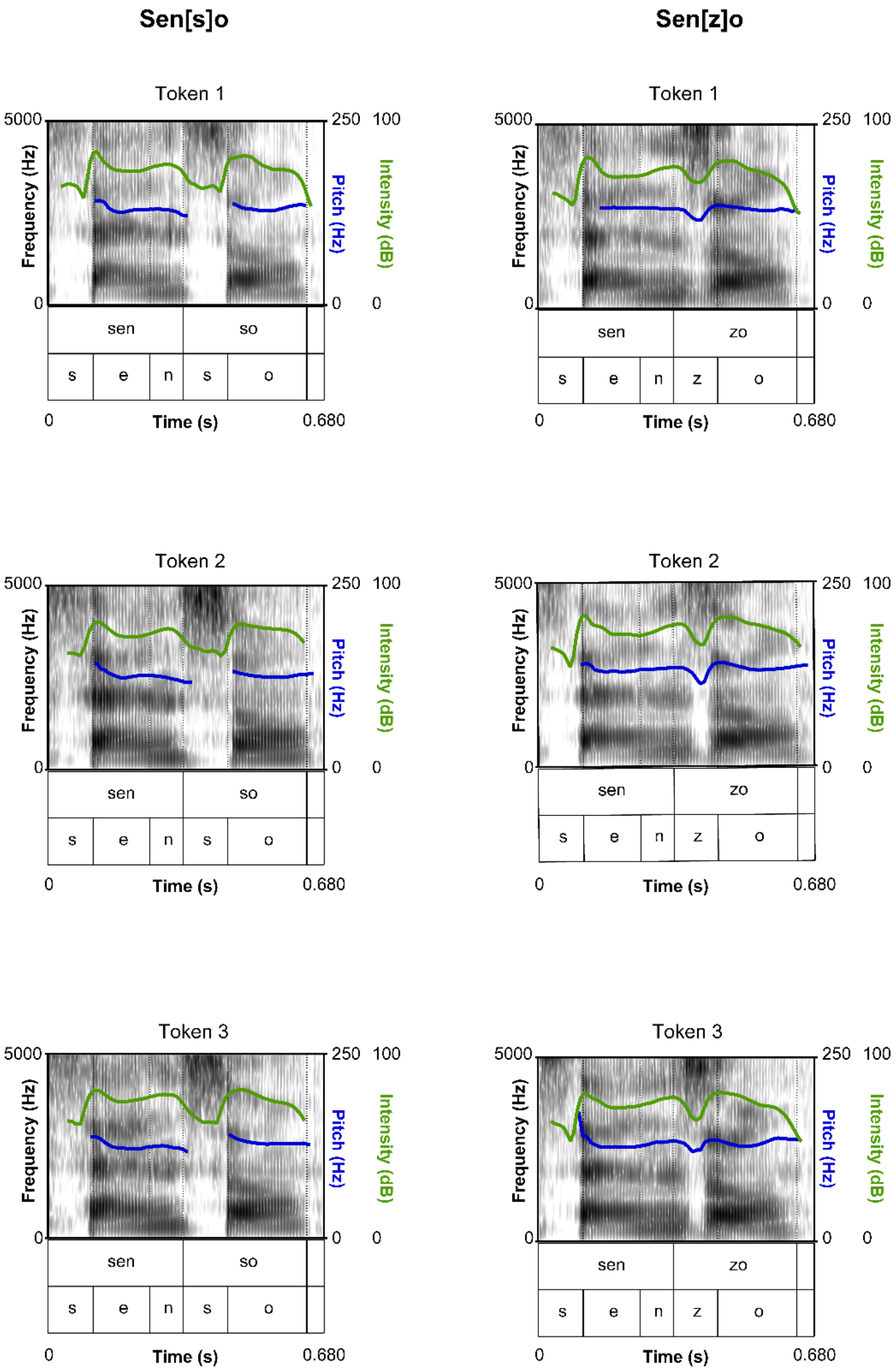

Figure S1. Spectrograms of the experimental stimuli with pitch profiles in blue (0-250 Hz) and intensity profiles in green (0-100 dB) on the y-axis. Right column shows the spectrograms for each of the three tokens for voiceless sen[s]o, while left column shows the spectrograms for each of the three tokens for voiced sen[z]o.

Items of The Sociolinguistic Questionnaire

1. How frequently do you speak Trentino dialect with your family members?

   ☐ Always
   ☐ Almost Always
   ☐ Sometimes
   ☐ Almost never
   ☐ Never

2. How frequently do you speak Trentino dialect with your friends?

   ☐ Always
   ☐ Almost Always
   ☐ Sometimes
   ☐ Almost never
   ☐ Never

3. How frequently do your family members speak Trentino dialect to you?

   ☐ Always
   ☐ Almost Always
   ☐ Sometimes
   ☐ Almost never
   ☐ Never

4. How frequently do your friends speak Trentino dialect to you?

   ☐ Always
   ☐ Almost Always
   ☐ Sometimes
   ☐ Almost never
   ☐ Never