# Behavioral and Electrophysiological Evidence for Early and Automatic Detection of Phonological Equivalence in Variable Speech Inputs

## Viktor Kharlamov[1,2], Kenneth Campbell[1], and Nina Kazanina[3]

## Abstract

■ Speech sounds are not always perceived in accordance with their acoustic–phonetic content. For example, an early and automatic process of perceptual repair, which ensures conformity of speech inputs to the listener's native language phonology, applies to individual input segments that do not exist in the native inventory or to sound sequences that are illicit according to the native phonotactic restrictions on sound co-occurrences. The present study with Russian and Canadian English speakers shows that listeners may perceive phonetically distinct and licit sound sequences as equivalent when the native language system provides robust evidence for mapping multiple phonetic forms onto a single phonological representation. In Russian, due to an optional but productive *t*-deletion process that affects /stn/ clusters, the surface forms [sn] and [stn] may

be phonologically equivalent and map to a single phonological form /stn/. In contrast, [sn] and [stn] clusters are usually phonologically distinct in (Canadian) English. Behavioral data from identification and discrimination tasks indicated that [sn] and [stn] clusters were more confusable for Russian than for English speakers. The EEG experiment employed an oddball paradigm with nonwords [asna] and [astna] used as the standard and deviant stimuli. A reliable mismatch negativity response was elicited approximately 100 msec postchange in the English group but not in the Russian group. These findings point to a perceptual repair mechanism that is engaged automatically at a prelexical level to ensure immediate encoding of speech inputs in phonological terms, which in turn enables efficient access to the meaning of a spoken utterance. ■

## INTRODUCTION

Psycholinguistic studies have consistently shown that language-specific experience can impose a bias on perception of individual speech sounds. For example, robust reparative effects of native phonology arise automatically at an early prelexical level when input segments do not have phonemic status in the native inventory (Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Winkler et al., 1999; Dehaene-Lambertz, 1997; Näätänen et al., 1997; Miyawaki et al., 1975; Goto, 1971). At the same time, the exact role of linguistic experience in the perception of sound *sequences* violating either optional or compulsory phonotactic restrictions of the native system is not at all clear, as it remains to be determined at what stage and under which conditions native phonology and lexical knowledge interact with the perceptual system during the processing of speech sound strings. A number of previous studies have shown that phonological influences on perception of segmental sequences can be language-specific, early (prelexical) and/or applying automatically to words and nonwords alike (e.g., Peperkamp & Dupoux, 2007; Beddor, Harnsberger, & Lindeman, 2002; Lahiri &

Marslen-Wilson, 2001; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Hallé, Segui, Frauenfelder, & Meunier, 1998), whereas in other studies, the observed effects appear to be language-universal, late (postlexical) and/or nonautomatic and affecting only a subset of the existing lexical items (Pitt, 2009; Mitterer, Yoneyama, & Ernestus, 2008; Gow & Im, 2004). In the present study, we investigate the role of native language phonology on the perception of sound sequences by examining the perception of optionally reduced consonant clusters in Russian. We report electrophysiological and behavioral evidence for early and automatic phonological effects that are highly language-specific and are also independent of the lexical status of the input string.

Studies dealing with the role of native phonology in the perception of segmental sequences initially examined issues related to the processing of phonotactically illegal consonant clusters. For example, Dehaene-Lambertz et al. (2000) and Dupoux et al. (1999) looked into the perception of illicit consonantal sequences by Japanese listeners whose native language places severe restrictions on the co-occurrence of consonants within a word. In particular, Japanese syllable structure requirements prohibit any sequence of two adjacent consonants (except when the first consonant is a nasal), thus rendering sequences such as [..gm..] or [..bz..] phonotactically illegal. Dupoux,

[1]University of Ottawa, Canada, [2]Queen's University, Canada, [3]University of Bristol, UK

Dehaene-Lambertz, and colleagues observed early and automatic perceptual epenthesis effects in the Japanese listeners who consistently misperceived inputs such as [igmo] as "igumo" in behavioral tasks. Electrophysiological data further showed that such misperception occurred at an early stage of processing within 150–200 msec from the onset of [m] in [igmo], as confirmed by the lack of mismatch negativity (MMN) in the corresponding time window in the data of the Japanese group (MMN to the deviant [igmo] vs. the standard [igumo] was present in the control group of French speakers for whom both sequences were licit). The authors argued that the misperception observed in the behavioral task and the lack of an MMN in Japanese listeners indicated a reparatory effect driven by the native language phonology. The illicit input [igmo] was presumably mapped onto /igumo/, the closest licit phonological representation available in Japanese. (An implicit assumption here is that a phonological match is needed to encode a speech input in memory.)

These results were taken as evidence for early, automatic, and language-specific effects of the native phonological system on the processing of sound sequences. Other studies have also reported similar early and automatic perceptual effects that arise in the case of illicit sound sequences (e.g., perception of word-initial /tl/ or /dl/ sequences by French speakers; Hallé et al., 1998). Such findings were in line with much of the literature on the processing of individual speech sounds; however, it remained to be determined whether a phonological account was, in fact, the most plausible analysis. In the case of Japanese, for example, it has been shown that listeners automatically segment speech input into moraic units (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993) and, therefore, the reparatory effect did not need to arise due to higher-order phonological factors but could instead emerge from a failure to parse the speech input into preexisting moraic templates. It also remained to be established whether such robust language-specific phonological influences were limited to perception of phonotactically illegal inputs or whether a more general and modulated perception of illicit and licit sound sequences alike was applied.

The question of whether phonology influences perception of *licit* sound sequences is important because regular phonological processes often result in a mismatch between underlying phonological representations and output forms. Consider, for example, nasal place assimilation in Dutch, whereby the place of articulation of a word-final nasal segment assimilates to that of the following consonant (e.g., the phonological sequence /n#b/ surfaces in speech as [m#b]; "#" indicates a word boundary). As a result of this process, Dutch listeners are faced with the task of recognizing inputs such as *tui*[m] as instances of the word *tuin* "garden" in the acoustic sequence *tiu*[m#b]*ank* "garden bench." Mitterer and Blomert (2003) used behavioral and MMN data to argue that, in such cases, Dutch listeners apply a prelexical compensatory mechanism

that enables them to recover the underlying form. This perceptual compensation took place only in viable contexts (i.e., when nasal place assimilation was licensed; e.g., in the [m#b] context but not in the [m#s] context). At the same time, similar compensatory effects were also found in the behavioral data from German speakers who did not have any direct experience with the Dutch words used as stimuli items but did have exposure to the same assimilation process in their native language, leading Mitterer and Blomert to conclude that compensation for assimilation can occur even in the absence of language-specific lexical knowledge. A similar claim that compensatory effects in perception rely on language-universal processing mechanisms can be found in Mitterer, Csépe, and Blomert (2006) and Gow and Im (2004). Mitterer et al. (2006), for example, examined how Hungarian and Dutch listeners process sound sequences affected by a liquid assimilation process found in Hungarian but not in Dutch (in Hungarian, word-medial /lr/ clusters surface as [rr]). Behavioral results revealed the presence of significant compensatory effects in both language groups; that is, Hungarian as well as Dutch listeners were worse at perceiving the intended targets in environments in which assimilation was viable, such as [lr] versus [rr], but were quite good at perceiving the same target segments in those environments in which assimilation was not viable, such as [ln] versus [rn]. These results were interpreted as evidence for a language-independent compensation mechanism that is driven by general auditory rather than phonological processing. Gow and Im (2004) explored how native and nonnative listeners perceived speech inputs affected by two types of assimilatory processes: (i) Hungarian voicing assimilation which causes word-final voiceless obstruents to become voiced when followed by a voiced obstruent (e.g., pronouncing /s#d/ as [z#d]), and (ii) Korean labial-to-velar place assimilation which turns word-final labial consonants into velars when followed by a velar segment (e.g., pronouncing /m#g/ as [ŋ#g]). For voicing assimilation, behavioral results revealed similar context-dependent compensatory effects in both Hungarian and English listeners, even though only the Hungarian group had previous direct linguistic experience with the assimilation in question. For place assimilation, no evidence for compensatory effects was found either in English listeners, who had no previous experience with the this type of assimilation, or Korean speakers, who did have exposure to the assimilated forms in their native language. Given that compensatory effects were found when listeners were unfamiliar with the specific assimilatory process *and* no context effects were present when listeners did have direct experience with the assimilation, Gow and Im concluded that language-specific experience cannot be the main factor in perceptual compensation for assimilatory phenomena.

A different conclusion, namely, that language-specific factors *can* affect the compensatory mechanism, has been reached by studies on the perception of lenited consonantal sequences in Dutch (Pitt, 2009; Mitterer et al., 2008; Mitterer

& Ernestus, 2006). In Dutch, the phonetic realization of a word-final /t/ varies from nonlenited (fully released) to lenited (released only partially, unreleased, or deleted), with lenition more likely to occur when /t/ is preceded by /s/ (as in /st/) than when preceded by /n/ (as in /nt/). Mitterer and Ernestus (2006) analyzed behavioral data from Dutch listeners and found that compensation for *t*-lenition was, in fact, stronger in the /st/ environment, which was highly viable for lenition, although listeners also appeared to rely on the presence of subphonemic cues and, to a lesser extent, on their lexical knowledge of whether or not there was a corresponding Dutch word containing a final /t/. They concluded that compensation for *t*-lenition must be dependent on higher-level linguistic knowledge. Mitterer et al. (2008) collected behavioral data from both Dutch and Japanese listeners as well as MMN data from Dutch participants to further examine the phenomenon of lenition in Dutch. The behavioral results suggested that the perceptual mechanism involved in the recognition of lenited /nt/ and /st/ sequences was sensitive to the language-specific viability of lenition in a given phonological context. On the other hand, Japanese listeners also showed some compensatory effects, thus suggesting a role for general auditory processing. Furthermore, the electrophysiological data from the Dutch participants revealed the presence of an MMN, which was due to the difference in responses to the standards and the deviants (e.g., [blantmuj] vs. [blanmuj]). MMN was present as early as 85–135 msec postchange, which Mitterer et al. associated with predominantly auditory processing. These differences were not apparent in a later 175–225 msec postchange interval, a time window the authors believed to be indicative of phonological processing. Mitterer et al. concluded that compensation for *t*-lenition involves a complex interplay of lexical, phonological, and general auditory factors.

Unlike the above-mentioned studies by Mitterer and colleagues that concentrated on a word-edge phenomenon, Pitt (2009) investigated a deletion process that occurs *word-medially*. Pitt explored how native speakers of English process reduced variants of existing word forms as well as newly learned nonword sequences. Speakers of American English often delete the /t/ in /nt/ clusters, but this type of reduction is dependent on the quality of the following vowel (e.g., *twen(t)y*, *cen(t)er*; cf. *contact*). Behavioral data for the existing English words showed that phonologically unviable reduced forms (e.g., *conact* as a reduced version of *contact*) were heard as nonwords, but phonologically viable forms (e.g., *cener* as a reduced form of *center*) were perceived similarly to their nonreduced counterparts. In contrast, newly learned nonword stimuli that exploited the same phonological context effect (e.g., *seny* as a reduced form of *senty*) did not show any effects of phonological inference unless the participants were informed indirectly that the two forms were related (in Pitt's study, participants "overheard" a conversation between two experimenters in which there was an indirect

reference to the fact that *seny* and *senty* were the same word form). Given that phonological generalization was not automatic and listeners required specific lexical knowledge in order to make an active connection between the full and the reduced forms, Pitt concluded that a strictly phonological account of the compensation mechanism could not be adequate. Because previous accounts of compensation were based primarily on word-final phenomena and sometimes found evidence for effects of phonological context but not lexicality (e.g., Mitterer & Blomert, 2003), Pitt proposed that lexical processes may play a greater role word-medially, whereas phonological processes may have more impact at word edges.

The present study is a further inquiry into compensation mechanisms involved in the recognition of sound sequences. Both behavioral and electrophysiological data were collected to determine the contribution of language-specific factors in the perception of reduced consonantal sequences in Russian. We examined how Russian listeners perceive two phonetically distinct nonsense auditory strings, [asna] and [astna], which may share a single phonological representation in their native language. In Russian, optional deletion of the medial segment in word-internal three-consonant clusters results in the phonological form /stn/ surfacing phonetically as either [stn] or [sn] both in casual and formal speech, yet Russian listeners have little difficulty in recognizing either variant as an instance of the same underlying form.[1] For example, the word *mestnyj* "local" can be pronounced with or without the /t/ as *me*[stn]*yj* or *me*[sn]*yj*, with the presence of an underlying /t/ reinforced paradigmatically via morphologically related forms in which the /t/ is prevocalic and does not delete (such as *mesto* "place, locality") and also by Russian orthography as "t" is mandatory in spelling. Given that Russian listeners frequently encounter [stn] and [sn] clusters and have a robust experience with the [sn]–[stn] alternation, they should be capable of matching either [sn] or [stn] to the phonological form /stn/. If native language phonology influences listeners' perception of licit consonantal sequences, the potential phonological equivalence of [stn] and [sn] clusters can be expected to result in higher confusability of the two forms.

To investigate the language specificity of the mechanism involved in coping with deleted medial /t/s, a control group of Canadian English speakers was also tested. In English, *t*-deletion is attested word-finally and in intervocalic /nt/ clusters (Guy, 1992, 1997; Neu, 1980; Zue & Laferriere, 1979) as well as in some three-consonantal sequences (e.g., /stl/; Raymond, Dautricourt, & Hume, 2006).[2] However, the phonetic sequence [sn] rarely represents the phonological string /stn/ in formal English speech (with some possible exceptions, such as *fa*[sn]*ess* as an output of *fastness*) and the word-medial [sn]–[stn] alternation is not as robust in English casual speech as it is in Russian casual speech. This cross-linguistic difference could be attributed, in part, to the lexical frequency of the English words containing the /stn/ sequence. Reduction of segments is

well known to be less likely in word forms of low usage frequency (Jurafsky, Bell, Fosler-Lussier, Girand, & Raymond, 1998), and all English words that have a medial /stn/ string are highly infrequent. A search of the SUBTLex frequency database of American English (Brysbaert & New, 2009), for example, would yield a total of only eight word forms containing a medial /stn/ cluster (e.g., *fastness*, *vastness, robustness*), all of which have a usage frequency of less than 1 item per million (henceforth, ipm). In comparison, the frequency dictionary of Russian compiled by Sharoff (2002) lists over 300 word forms that contain /stn/ word-medially, with the average frequency exceeding 6 ipm and the highest individual frequency value of above 160 ipm for the word *izvestno* "known." Thus, unlike Russian with its high frequency of /stn/ word forms, high productivity of the *t*-deletion process in the /stn/ cluster and no /stn/ words in which *t*-deletion is blocked, English offers only limited exposure to the [sn]–[stn] alternation in word-medial environments and does not provide a solid basis for treating two phonetic forms [sn] and [stn] as phonologically equivalent (especially not in the case of a more formal speech style). Note also that although in English there are contexts in which *t*-deletion is quite common, it still has a restricted scope of application (e.g., compare the viability of *t*-deletion in the /nt/ cluster in *plenty* with the unviability of deletion in the same /nt/ cluster in *pontiff*). As such, English speakers must know that *t*-deletion cannot be freely generalized to any phonological context.[3]

Therefore, if the phonological regularity that develops as a result of robust exposure to the [sn]–[stn] alternation in one's native system can exert early and automatic influences on perception, we should expect to see significant differences in processing between the two language groups. In particular, the stimulus [asna] presented to Russian listeners can be expected to invoke an underlying representation /asna/ as well as /astna/, as both forms can result in the output [asna]. In contrast, English listeners should match the input [asna] primarily onto the phonological representation /asna/. As such, significantly lower accuracy scores and/or longer reaction times when identifying the [asna] sequence or discriminating [asna] from [astna] should be obtained for Russian as compared to English listeners. Likewise, between-language differences can also be expected in the event-related potentials (ERPs) recorded while participants listen passively to multiple repetitions of the items [astna] and [asna] presented in an oddball paradigm. In such a paradigm, many tokens of one stimulus type, the so-called standard, are interspersed with a few tokens of the other stimulus type, known as the "deviant," and detection of an acoustic difference between deviants and standards yields an MMN response (Näätänen, Paavilainen, Riine, & Alho, 2007; Näätänen et al., 1997) approximately 150–250 msec after stimulus onset.[4] If memory traces formed during passive listening reflect perceptual differences emerging due to language-specific phonological processes not only at the level of individual segments, as it has been previously suggested in the literature (among others, Näätänen, 2001; Phillips et al., 2000; Näätänen et al., 1997), but also at the level of segmental strings, then significant differences should be found in the MMN components of the two language groups, with MMN present in English but not Russian participants in the [asna]–[astna] conditions. The consonant cluster *skn* that exists in both Russian (*vypu*[skn]*ik* "graduate") and English (*bri*[skn]*ess*) and is not subject to deletion in either language (i.e., [skn] cannot be realized as [sn]) was used as a control, with no significant differences in processing expected between the two language groups in the [asna]–[askna] conditions.

## METHODS

### Participants

Participants were adult native speakers of Russian and English. Russian speakers were recruited in Ottawa, Canada (*n* = 12, 5 men; age range 18–39 years) and in Perm, Russia (*n* = 24, 11 men; age range 18–36 years). The Russian participants recruited in Ottawa took part in an identification task and in the EEG recording. They indicated having intermediate to advanced proficiency in English and reported using Russian on a daily basis. Before coming to Canada, they did their primary and secondary schooling in Russian. The participants from Perm had limited exposure to languages other than Russian and were tested in both identification and discrimination experiments to ensure that the performance of the Russian speakers recruited in Canada was not critically affected by their exposure to English. The English group was recruited in Ottawa, Canada. Twelve participants (6 men; age range 18–36 years) took part in the identification task and in the EEG experiment and 20 participants (10 men; age range 18–38 years) performed the discrimination task. All participants were right-handed, had no previous history of hearing problems or language disorders, and gave informed consent for participation.

### Stimuli

The stimuli were three bisyllabic nonwords that differed in the medial consonant cluster: [asna], [astna], and [askna]. The tokens were created using PRAAT (Boersma, 2001) by combining the sequences [as] and [na], [tna] or [kna] (each cut out from a larger context) produced by a female speaker in Russian. The resultant medial [stn] and [skn] sequences contained a period of frication noise of the [s], followed by an oral closure and a brief release burst of the [t]/[k], and a nasal closure of the [n]. The items were physically identical up until the third segment (the [n], the [t], or the [k]), which always started at approximately 325 msec after the onset of the stimulus. For each stimulus type, three tokens were created on the basis of three different natural recordings by the same speaker. Introducing such acoustic variability was intended to stimulate more abstract, phonological processing of the input

forms. All tokens were acceptable to both Russian and English speakers as instances of *asna*, *astna*, and *askna*. The [asna] tokens were 470 msec long, whereas the [astna] and [askna] tokens were 550 msec long. The stimuli were presented binaurally via headphones (behavioral tasks) or insert earphones (Etymotic, ER-3A) at 75 dB peak SPL (EEG recordings).

## Behavioral Paradigm

In the identification task, participants heard a nonword and were asked to identify it by pressing one of two designated buttons on a computer keyboard. There were two experimental blocks with 60 trials per block. In the *t-Ø* block, participants heard either [asna] ($n = 30$) or [astna] ($n = 30$) and had to choose between two buttons labeled as *asna* and *astna*. In the *k-Ø* block, participants heard either [asna] ($n = 30$) or [askna] ($n = 30$) and had to choose between the buttons labeled as *asna* and *askna*. The stimuli were randomized within each block and the order of blocks was counterbalanced across subjects. In the discrimination task, participants heard one pair of tokens at a time (e.g., [asna]–[astna]) and were asked to indicate whether the two tokens were the same or different. There were 144 trials in total (72 "same" and 72 "different"). The *same* pairs always involved acoustically different tokens of the same nonword. In both tasks, accuracy and response times were recorded for analysis.

## EEG Acquisition and Analyses

To avoid creating any attentional bias in the electrophysiological study, the EEG recording always preceded behavioral testing. Continuous EEG recordings were made at a sampling rate of 256 Hz using 12 electrodes placed according to the International 10–20 System (Fz, Cz, Pz, Oz, F3, F4, P3, P4, T7, T8, M1, M2) and a nose tip reference. Electrooculogram (EOG) was coregistered to control for eye movements and blinks. A vertical EOG was recorded using electrodes placed at the supra- and infraorbital ridges. A horizontal EOG was recorded using electrodes placed at the outer canthus of each eye. The impedance of all electrodes was below 5 kΩ. The EEG and EOG signals were filtered on-line using a 35-Hz low-pass filter and a 1-sec time constant.

The stimuli were presented in four experimental blocks using an oddball paradigm. Each block contained two stimulus categories: the standard (850 repetitions) and the deviant (150 repetitions). The stimuli were presented in a pseudorandomized order with at least five standards separating any two adjacent deviants. The stimulus onset asynchrony interval varied randomly from 1000 to 1100 msec. In the first block, the standard stimulus was [asna] and the deviant stimulus was [astna]. In the second block, the standard [asna] was coupled with the deviant [askna]. Standards and deviants were reversed in the two remaining blocks (e.g., [asna] was now the deviant and [astna] the standard). The order of blocks was randomized across participants. During the recording, participants were seated in a sound-attenuated room and were watching a silent movie of their choice.

The continuous EEG data were divided into 1000-msec-long epochs starting at 100 msec prior to the onset of the auditory stimulus. The epochs with either EEG or EOG activity exceeding ±100 μV were excluded automatically. The recordings were also inspected visually for any additional artifacts. The remaining data were averaged and baseline-corrected by using a 100-msec prestimulus interval.

## RESULTS

### Behavioral Results

For the identification task, mean error rates and response times from 12 English and 12 Russian participants tested in Canada (Table 1) were entered as dependent variables in repeated measures ANOVAs with Contrast (t–Ø, k–Ø) as a within-subject factor and Language (Russian, English) as a between-subject factor. Although accuracy rates were high in all conditions in both language groups, there was a statistically robust main effect of Language [$F(1, 22) = 6.12$, $p = .022$] and a Language × Contrast interaction [$F(1, 22) = 8.12$, $p = .009$]. Resolving the interaction within each language revealed a significant main effect of Contrast in the Russian group only [Russian: $F(1, 11) = 8.05$, $p = .016$; English: $F < 1$]. The Russian group made significantly more errors to the t–Ø contrast than to the k–Ø contrast; the English group did not. In the RTs, both the main effect of Contrast [$F(1, 22) = 7.73$, $p = .011$] and the Language × Contrast interaction [$F(1, 22) = 4.42$, $p = .047$] were significant. Separate ANOVAs for each language group revealed a significant main effect of Contrast in the Russian participants only [Russian: $F(1, 11) = 20.72$, $p = .001$; English: $F(1, 11) = 0.161$, $p = .696$], due to significantly longer RTs in the t–Ø than in the k–Ø condition, but only for the Russian group. The results from the Russian group tested in Russia were consistent with the results from the Russian group tested in Canada: A significant effect of Contrast was observed both in the error rates [$F(1, 23) = 18.37$, $p < .001$] and the RTs [$F(1, 23) = 8.77$, $p = .007$]. Thus, the results of the identification task demonstrate that Russian participants were significantly less accurate and slower in identifying the nonword [asna] as *asna* in those conditions where the competitor choice involved the [stn] cluster (i.e., *asna* vs. *astna*) than when it did not (i.e., *asna* vs. *askna*). The English group did not show any significant differences between the conditions and had higher overall accuracy.

For the discrimination task, mean error rates and response times from 20 English and 24 Russian participants tested in Russia (Table 1) were entered as dependent variables in repeated measures ANOVAs with contrast (t–Ø, k–Ø) and order (addition of a consonant, removal of a

**Table 1.** Results of the Identification and Discrimination Tasks for the Russian (from Perm, Russia) and English (from Ottawa, Canada) Speakers: Mean Error Rates (%), Mean Response Times (msec), and Corresponding Difference Scores

| Language Group | Identification | | | Discrimination | | |
|---|---|---|---|---|---|---|
| | t | k | Diff | t–Ø | k–Ø | Diff |
| *Russian* | | | | | | |
| Error rate | 7.1 [3.9] | 3.2 [1.5] | 3.9*** [2.4**] | 25.5 | 10.2 | 15.3*** |
| RT | 766 [766] | 742 [737] | 24* [29*] | 1039 | 984 | 55*** |
| *English* | | | | | | |
| Error rate | 0.7 | 1.1 | −0.4 | 5.37 | 5.08 | 0.29 |
| RT | 752 | 756 | −4 | 1152 | 1141 | 11 |

The results of the identification task for the Russian group tested in Ottawa are given for comparison in square brackets.

*p < .01.

**p < .05.

***p < .001.

consonant) as within-subject Factors and Language (Russian, English) as a between-subject factor. The factor Order was introduced to examine potential effects related to the order of the stimuli within each pair (e.g., hearing [asna] followed by [astna] involves addition of the [t] in the second item of the pair, whereas [astna]–[asna] pairs involve removal of the [t]). The error rate data revealed a statistically robust main effect of Language [$F(1, 42) = 18.3, p < .001$], a significant main effect of Contrast [$F(1, 42) = 17.22, p < .001$], and a significant Language × Contrast interaction [$F(1, 42) = 15.94, p < .001$]. The effect of Order was only marginally significant [$F(1, 42) = 3.34, p = .071$] and did not interact with any other factor (all $Fs < 1$, all $ps > .1$). Separate ANOVAs for each language revealed a significant main effect of Contrast in the Russian group only [Russian: $F(1, 23) = 22.14, p < .001$; English: $F(1, 19) = 0.05, p = .822$]. The Russian group made significantly more errors to the t–Ø contrast than to the k–Ø contrast; the English group did not. The effect of Order was marginally significant in the English group [$F(1, 19) = 4.0, p = .060$], who showed a tendency for higher error rates in the case of segment removal. The Order factor was not significant in the Russian group and neither group has a significant Contrast × Order interaction (all $Fs < 1$). The RT data revealed a significant main effect of Contrast on the RTs of Russian but not English participants [Russian: $F(1, 22) = 20.03, p < .001$; English: $F(1, 19) = 1.19, p = .289$]. The Russian group had significantly slower reaction times in the case of the t–Ø contrast than the k–Ø contrast; the English group did not show a significant difference in RTs across the two contrasts. The Order factor was not significant and did not interact with contrast in either language group (all $Fs < 2$, all $ps > .1$). Together, these results demonstrate that Russian, but not English, participants made significantly more discrimination errors and had slower response times when [asna] was compared

to [astna] than when it was compared to [askna]. Cumulatively, the results of the identification and discrimination tasks highlight confusability of [sn] versus [stn] clusters for Russian speakers, which is specific to a given phonological context (hence, not observed with [sn] vs. [skn] clusters) and linguistic background (hence, not observed with English speakers).

## EEG Results

For each language group, ERP data from five electrode sites were analyzed using repeated measures ANOVAs with factors Category (standard, deviant), Contrast (t–Ø, k–Ø), Order (addition of a consonant, removal of a consonant), and Electrode (Fz, Cz, Pz, F3, F4). Greenhouse–Geisser correction was applied wherever appropriate (in such cases, the corrected $p$ value is reported alongside the original number of degrees of freedom). The averaged ERPs to each contrast from both language groups are shown in Figure 1.

Mean ERPs to the standard and the deviant categories were analyzed in nine 100-msec time intervals starting with the stimulus onset. There were no significant main effects or interactions in the ERPs of either language group prior to 400 msec poststimulus onset with the exception of the main effect of Electrode, which was present in the majority of analysis windows in both language groups. Amplitude differences across different electrode sites are expected; they reflect the largely sensory, exogenous influences on the processing of both the standards and the deviants. Further, the same identical physical stimulus served as standards and deviants in the different blocks; early sensory processing should therefore be identical for the standards and deviants. It is therefore the difference in largely nonsensory, endogenous processing when

**Figure 1.** Averaged ERPs to the standard (black) and the deviant (gray) categories from the English (A–B) and Russian (C–D) participants at the Fz electrode. Each solid line represents averaged data from two blocks (e.g., for the k–Ø contrast, the black line is an average of the standard *askna* from the *askna–asna* block and the standard *asna* from the *asna–askna* block; for the same k–Ø contrast, the gray line is an average of the deviants *asna* from the *askna–asna* block and *askna* from the *asna–askna* block). "0 msec" signals the onset of the auditory stimulus ([asna], [astna] or [askna]); a dotted line marks the onset of the acoustic difference between the standard and the deviant (~325 msec poststimulus onset). Shading marks time windows of (marginally) significant differences ($p \leq .05$) based on pairwise comparisons of the standard and the deviant.
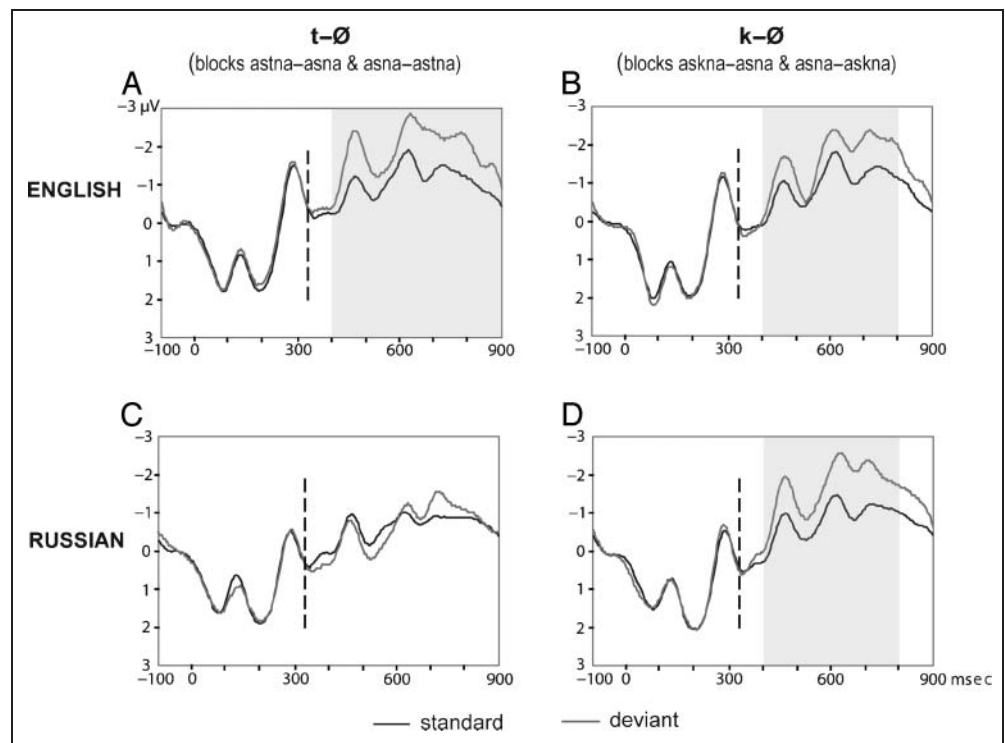


the same stimulus served as a standard and as a deviant that was compared. Exogenous influences will not be discussed further.

### Russian

A main effect of Contrast was significant in the 400–500 msec time window [$F(1, 11) = 9.58, p = .010$] and in the following three time windows up to 800 msec poststimulus onset (all $Fs > 6.30, ps < .05$). Crucially, in the 400–500 msec time window, there was also a significant Contrast × Category interaction [$F(1, 11) = 4.98, p = .047$], which remained significant in the two subsequent windows up to 700 msec poststimulus (all $Fs > 5.14, ps < .05$). The main effects of Category or Order, or their interaction, did not reach significance in any time window. Planned pairwise comparisons of the standard and the deviant responses (two-tailed paired $t$ tests) were run on the Fz data, where the MMN was largest in all time windows. For pairwise comparisons, a token that represented the standard category was compared to itself when it represented the deviant category in order to avoid differences that are sensory/acoustics-based. For example, the *askna* standard in the *askna–asna* block was compared to the *askna* deviant in the reversed *asna–askna* block. For the k–Ø contrast, the ERP response to the deviant revealed a significantly greater negativity in the windows from 400 up until 800 msec (all $|t|s > 2.50$, $ps < .05$). By contrast, there were no significant differences between the standard and the deviant waveforms

for the t–Ø contrast in any of the time windows in the 400–900 msec interval (all $ts < 1.6, ps > .1$).

### English

Neither a main effect of Contrast nor a Contrast × Category interaction was found in any of the time windows. A significant main effect of Category [$F(1, 11) = 13.23, p = .004$] was found in the 400–500 msec window; ERPs were more negative to the deviant category as compared to the standard category and remained significant in all time windows up to 900 msec poststimulus (all $Fs > 5.60, ps < .05$). The Order factor was not significant in any time window. A significant Order × Category interaction was found only in the 500–600 msec window [$F(1, 11) = 6.68, p = .025$], and resolving this interaction showed that the effect of Category was significant only in the "removal" blocks in which the deviant stimulus had one segment less than the standard ($p = .006$, Bonferroni-corrected). Planned pairwise comparisons at the Fz electrode site revealed that, for the t–Ø contrast, the response to the deviant showed a significantly greater negativity than the response to the standard starting with 400 msec and up until 900 msec (all $|t|s > 2.97, ps < .05$). For the k–Ø contrast, the deviant waveform showed greater negativity than the standard waveform at a marginally significant level in the 400–500 msec interval [$t(11) = -2.17, p = .052$] and at a significant level in the following three time windows from 500 msec until 800 msec (all $|t|s > 2.27, ps < .05$).

In sum, in the case of the t–Ø contrast for which the standard and the deviant could be perceived as phonologically equivalent in Russian but not in English, an MMN-like component was found in the English group data starting approximately 100 msec after the onset of the acoustic difference between the standard and the deviant (Figure 1A); such an early MMN was absent from the Russian ERPs (Figure 1C). In the case of the k–Ø contrast for which the standard and the deviant had distinct phonological representations in Russian and English alike, the MMN was found starting about 100 msec after the onset of the acoustic difference in both language groups (Figure 1B and D).

## DISCUSSION

In the present study, we used behavioral and electrophysiological techniques to investigate Russian listeners' processing of licit consonantal clusters [sn] and [stn] that are phonetically distinct but phonologically equivalent in their native system. English listeners whose native language did not provide robust evidence for phonological equivalence of the sequences in question served as a comparison group. The results of the behavioral and ERP experiments confirmed the existence of early, automatic, and language-specific phonological influences on perception. In the behavioral tasks, Russian participants had higher error rates and slower response times in the conditions involving the t–Ø contrast than in those involving the k–Ø contrast (regardless of whether they lived in Canada and had exposure to English, or in Russia and had only minimal exposure to English), whereas the English group showed a similar performance on the two contrasts. In the ERP experiment, the morphology and distribution of the observed effect was consistent with the descriptions of the typically reported MMN (e.g., Näätänen et al., 1997, 2007; Näätänen, 1995). The Russian group had an early and robust MMN to the k–Ø contrast starting within about 100 msec postchange, but not to the t–Ø contrast. Conversely, with the English group, an MMN was present in the averaged waveforms for both the k–Ø and the t–Ø contrasts, also starting within approximately 100 msec postchange. The observed pattern of responses demonstrates that, due to their perceived phonological equivalence, phonetically distinct items such as [asna] and [astna] were treated as analogous by Russian listeners already at an early stage of processing.

Although our standard and deviant categories contained multiple instances of the same nonword (*asna*, *astna*, or *askna*) in order to induce more abstract processing, within-category acoustic variation was smaller than between-category variation. For example, the three instances of the nonword *asna* had an essentially identical duration and were acoustically more similar to one another than to any instance of the other two nonwords *astna* or *askna*. Yet, acoustic–phonetic differences alone were insufficient to elicit an MMN between *asna* and *astna* in the Russian group. In short, acoustic–phonetic considerations were not a sufficient basis for segregating a continuous stream of nonwords presented in an oddball paradigm into the envisioned standard and deviant categories, unless such segregation was also supported by phonology. Hence, our findings confirm that native language phonology has a profound effect on the sensory memory traces for speech stimuli and, consequently, on early automatic change-detection as reflected by the MMN. More generally, this interpretation supports the idea that speech input is immediately encoded in phonological terms, thus expediting access to the utterance meaning (Kazanina, Phillips, & Idsardi, 2006; Whalen et al., 2006; Dehaene-Lambertz & Gliga, 2004; Whalen & Liberman, 1987). It is also notable that the timing of phonological influences on perception of sound sequences is comparable with the timing of phonological influences on perception of individual segments, with both types of influences having an impact on the MMN within about 100 msec postchange. This finding provides support to the position that speech inputs are processed using short (segment-sized) and long (syllable-sized) time windows simultaneously (Poeppel, 2003; Zatorre, Evans, Meyer, & Gjedde, 1992).

The robust effects of native language phonology reported in the present article are in line with earlier ERP studies on perception of individual sounds (Kazanina et al., 2006; Eulitz & Lahiri, 2004; Sharma & Dorman, 1999, 2000; Dehaene-Lambertz, 1997; Näätänen et al., 1997) and complement previous electrophysiological findings on the processing of illicit sound sequences (Dehaene-Lambertz et al., 2000). As noted in the Introduction, earlier demonstrations of phonological influences on perception of illicit sound combinations could receive a competing nonphonological explanation in terms of segmentation into moraic or syllabic units. By examining phonetically licit clusters, the present study makes it possible to distinguish between the two accounts. Russian (but not English) listeners fail to differentiate between two distinct and highly familiar consonant clusters [sn] and [stn] during the early stages of processing, an outcome that cannot be explained in terms of speech segmentation requirements. Instead, perceptual similarity of these forms must be due to their phonological equivalence in Russian. Previously, Mitterer and Blomert (2003) examined phonological influences on the perception of licit phonetic sequences in Dutch. They found that the presence of optional nasal assimilation in Dutch forced Dutch listeners to perceive assimilated (*tui*[m#b]*ank*) and nonassimilated (*tui*[n#b]*ank*) sequences as equivalent. However, their ERP experiment did not contain a cross-linguistic comparison group, and thus, it remained ambiguous whether the perceptual assimilation found in Dutch was indeed due to a language-specific phonological process or due to low-level and universal acoustic/phonetic reasons. In the current study, a differential response to the t–Ø contrast by the Russian and English participants unambiguously points toward a language-specific nature of the observed early perceptual phenomenon. This finding is largely in line with other studies that examined

lenition processes and reported the existence of at least some language-specific effects (Pitt, 2009; Mitterer et al., 2008; Mitterer & Ernestus, 2006).

The present results further suggest that the perceptual compensation mechanism for *t*-deletion is context-specific rather than a general filter for dealing with lenited /t/s anywhere in the language. Namely, the readiness of Russian speakers to treat the two phonetic forms [asna] and [astna] that differ by the presence of the consonant [t] as equivalent is revealing considering that the [t] cannot be disregarded in other environments (e.g., the exact same consonant can be a suffix which differentiates the first-person singular and the third-person plural forms of Russian verbs, as in *poju*[Ø] "I sing" vs. *poju*[t] "they sing"). This fact highlights that the observed insensitivity to the t–Ø contrast in Russian listeners is context-dependent and is contingent upon a specific phonological process that optionally deletes the *t* between the *s* and the *n* in word-internal positions. This knowledge is available to Russian speakers from an early stage of processing (as indicated by the absence of an MMN to the t–Ø contrast in the 400–500 and 500–600 msec time windows) and applies even in the absence of an active linguistic task. Similarly, the fact that English speakers treated [sn] and [stn] as distinct suggests that their experience with the *t*-deletion process in other phonological environments, the general knowledge of the /t/ being more likely to delete than, for example, the /k/ in any consonant cluster of a given size (which is known to be the case in many languages, including English), and/or the infrequent experience with the [sn]–[stn] alternation do not lead to the development of a robust compensation mechanism that deals with *t*-deletion in all possible contexts.

The discussion above suggests that in order to establish a phonological generalization and a compensatory procedure for *t*-deletion in /stn/ clusters, listeners require a direct experience with the specific segmental context (/stn/). We propose that such a conservative compensation mechanism is due, at least in part, to the nature of the phonological deletion process. Were the compensation mechanism nonconservative and easily extendable to new phonological environments beyond /stn/, this could turn out to be highly problematic: If deletion of a segment were suspected in all instances, attempts to compensate for it would lead to enormous processing costs. Hence, there must be well-defined limits as to when *t*-deletion is acceptable in a given language (in which case it may be phonologized) and, consequently, when compensation for deletion should be applied. Interestingly, compensation mechanisms appear to be rather conservative even in the case of phonological processes that do not involve deletion (Peperkamp & Dupoux, 2007; Maye & Gerken, 2001). In particular, when adults are trained on a phonological regularity that does not exist in their native language, they can typically extend it to novel, untrained items but not to other phonological categories. Peperkamp and Dupoux (2007), for example, presented French speakers with an artificial language in which fricatives (or stops) underwent allophonic voicing, with fricatives (or stops) always being voiced in an intervocalic position and voiceless elsewhere, a regularity that does not exist in French. French participants learned the voicing rule and could compensate for it when presented with novel items that were not part of the training set, but they did not extend the generalization to consonants with a dental place of articulation (only labials, palatals, and velars were used in the training).

Our finding that the compensation mechanism for *t*-deletion in Russian is not limited to existing words but also extends to nonwords contrasts with Pitt's (2009) finding for English *t*-deletion in /nt/ clusters. Pitt demonstrated that American English speakers could compensate for deleted /t/s with existing words but not with nonwords. One possible reason for the lack of generalization to nonwords concerns the phonological structure of *senty* and *surnty*, the nonword items used in Pitt's study. In English, *t*-deletion does not necessarily affect all existing words containing the /nt/ cluster followed by an unstressed vowel. For example, whereas deletion is common in American usage of words such as *twenty* and *plenty*, the /t/ is almost always retained in the same segmental environment in words such as *pontiff*, which Pitt used in the condition where *t*-deletion was not expected (Appendix A.1., Pitt, 2009). Thus, even though *senty* and *surnty* were similar to *twenty* and *plenty*, in which deletion occurs very frequently in American English, their phonological structure was also quite similar to that of *pontiff, plaintiff*, and other such word forms in which deletion is, at best, rare. Therefore, the reason why Russian speakers generalized the *t*-deletion rule to nonwords, whereas English speakers did not, may be due to the fact that the /stn/ environment in Russian yields the [stn]–[sn] alternation more consistently than the /nt/ environment in English results in the [nt]–[n] alternation. More research is needed to evaluate this hypothesis and, more generally, to identify the conditions under which phonological regularity becomes sufficiently robust to operate automatically and without recourse to lexical information. Furthermore, as suggested by an anonymous reviewer, additional investigation is required to determine whether the absence of compensation for *t*-deletion in English speakers is largely due to the formal style of the stimuli items used in the present study. Lenited forms may be more expected (and thus, more strongly compensated for) in a casual speech style for English speakers; that is, compensation mechanisms may vary with the speech style.

Finally, the fact that Russian listeners treated the stimuli [asna] and [astna] as phonologically equivalent (as inferred from the absence of an MMN for the t–Ø condition) even though the two phonetic forms were not identical with respect to their phonological mappings suggests that the underlying representation /asna/ was dispreferred for the phonetic form [asna] by the Russian listeners. Recall that the phonetic form [asna] may be mapped to the phonological form /astna/ or /asna/ (using an example from existing Russian words, [sn] can map to either /stn/ as in *mestnyj*

"local" or to /sn/ as in *krasnyj* "red"). The observed preference for the competing phonological representation /astna/ could be due to the high productivity of the *t*-deletion rule in Russian as well as the presence of a concomitant stimulus [astna], which is obligatorily mapped onto /astna/ and, therefore, boosts the salience of the latter phonological form.

To conclude, the results of our investigation argue that early, automatic, and language-specific phonological influences on speech perception are not limited to phonemic inventories or reparatory effects targeting multisegmental phonetic sequences which are illicit in the listener's native language. Instead, listeners may treat licit and frequently-occurring inputs as equivalent even when they are clearly distinct at a phonetic level, as long as the native system provides robust evidence for phonological equivalence of the forms in question and the mapping process is not inhibited by higher-level knowledge. This further highlights the role of native language phonology as an early and automatic filter for speech inputs. Phonological representations are stable long-term memory abstractions that are critical for retrieving utterance meaning from speech with its substantial acoustic variability. To this end, the idea of primacy of phonological influences goes hand-in-hand with a more general view of information transfer as the main goal of language processing.

## Notes

1. The norm of the so-called literary pronunciation, which has been reiterated in many prescriptive Russian grammars and dictionaries, is for /t/ to be deleted completely in the output either without exceptions or with a possible exception of careful speech and extremely infrequent/bookish words (among others, Bogdanova, 2001; Gorbachevich, 2000; Avanesov, 1972). However, there is substantial empirical evidence suggesting that *t*-deletion is, in fact, optional in the majority of words containing a medial /stn/ sequence and that both fully realized ([stn]) and simplified ([sn]) output forms are common/acceptable in a wide range of words extending beyond literary-only terms or hyperarticulated speech (Kasatkin, 2006; Derwing & Priestly, 1980; Panov, 1967). The tendency to maintain [t] in production has also been described as a sound change in progress, with younger generations being more likely than the generations of their parents/grandparents to "pronounce the clusters much more as they are spelled" (Derwing & Priestly, 1980, p. 41).
2. Raymond et al. (2006) report the deletion rates of 50% and 28.6% for coronal plosives in the coda position preceded by /s/ and followed by /n/, respectively, without specifying the exact segmental makeup of the clusters in question. Our own search of the subset of the Buckeye corpus upon which Raymond and colleagues based their counts revealed no instances of /stn/ words, suggesting that the behavior of /t/ in the /stn/ cluster cannot be inferred from their data.
3. In order to verify that *t*-deletion in /stn/ clusters is a robust yet optional phonological process in modern Russian but not in English, we recorded two Russian and two English adult male speakers reading a list of /stn/ words embedded in a carrier sentence. The speakers were not aware of the exact goal of the study and the experimental list contained numerous filler items. In Russian, a medial [t] was present in 37 out of a total of 72 tokens, for a deletion rate of 51.4%. Notably, in many cases where a [t] was apparent in the spectrogram, the realization of the segment was quite weak and difficult to identify on the basis of auditory analysis alone. In English, a medial [t] was observed in 59 out of a total of 68 tokens, resulting in a *t*-deletion rate of 13.2%. Although more extensive independent verification is clearly required in future research, these results support the assumption used in the present investigation that the *t*-deletion process in /stn/ clusters is much more robust in Russian than in English.
4. Although the MMN component was originally conceived as an index of the detection of physical change, it can be elicited by any sound violating the regularity in an acoustic pattern (Näätänen & Winkler, 1999). A number of studies have also now demonstrated that the MMN is elicited automatically, relatively independent of attention and ongoing task demands (Muller-Gass, Stelmack, & Campbell, 2006; Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Näätänen, 1990).

## REFERENCES

Avanesov, R. (1972). *Russkoe literaturnoe proinoshenie* [Russian literary pronunciation]. Moscow: Prosveshchenie.

Beddor, P., Harnsberger, J., & Lindeman, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics, 30,* 591–627.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Bogdanova, N. (2001). *Proiznoshenije i transkriptsija* [Pronunciation and transcription]. St. Petersburg: Publishing House of the Philological Faculty of St. Petersburg State University.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behaviour Research Methods, 41,* 977–990.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33,* 824–844.

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

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12,* 635–647.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience, 16,* 1375–1387.

Derwing, B., & Priestly, T. (1980). *Reading rules for Russian.* Columbus, OH: Slavica Publishers.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1568–1578.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16,* 577–583.

Gorbachevich, K. (2000). *Slovar trudnostej proiznoshenija i udarenija v sovremennom russkom jazyke* [Dictionary of difficulties of pronunciation and stress placement in Modern Russian]. St. Petersburg: Norint.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9,* 317–323.

Gow, D., & Im, A. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51,* 279–296.

Guion, S., Flege, J., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107,* 2711–2724.

Guy, G. (1992). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change, 3,* 1–22.

Guy, G. (1997). Violable is variable: Optimality theory and linguistic variation. *Language Variation and Change, 9,* 333–347.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 592–608.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. (1998). Reduction of English function words in Switchboard. In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing, ICSLP-98, 30th November–4th December, 1998, Sydney, Australia* (Vol. 7, pp. 3111–3114). Canberra: Australasian Speech Science and Technology Association.

Kasatkin, L. (2006). *Sovremennij russkij jazyk. Fonetika.* [Modern Russian language. Phonetics.] Moscow: Academia.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 11381–11386.

Lahiri, A., & Marslen-Wilson, W. D. (2001). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38,* 245–294.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 480–490). Somerville, MA: Cascadilla Press.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65,* 956–969.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59,* 1305–1334.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34,* 73–103.

Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language, 59,* 133–152.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Muller-Gass, A., Stelmack, R. M., & Campbell, K. B. (2006). The effect of visual task difficulty and attentional direction on the detection of acoustic change as indexed by the mismatch negativity. *Brain Research, 1078,* 112–130.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13,* 201–288.

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

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38,* 1–21.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

Näätänen, R., Paavilainen, P., Riine, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* 2544–2590.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125,* 826–859.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English: A statistical analysis. In W. Labov (Ed.), *Locating language in time and space* (pp. 37–54). New York: Academic Press.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 258–278.

Panov, M. (1967). *Russkaja fonetika* [Russian phonetics]. Moscow: Prosveshchenie.

Peperkamp, S., & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. Hualde (Eds.), *Laboratory phonology* (Vol. 9, pp. 315–338). Berlin: Mouton de Gruyter.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., McGinnis, M., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12,* 1038–1055.

Pitt, M. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61,* 19–36.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric" sampling in time. *Speech Communication, 41,* 245–255.

Raymond, W., Dautricourt, R., & Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change, 18,* 55–97.

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

Sharma, A., & Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107,* 2697–2703.

Sharoff, S. (2002). *The frequency dictionary for Russian. Version 2*. Online resource. Retrieved January 19, 2009, from www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top–down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research, 13,* 393–405.

Whalen, D., Benson, R., Richardson, M., Swainson, B., Clark, V., Lai, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America, 119,* 575–581.

Whalen, D., & Liberman, A. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169–171.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357–369.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256,* 846–849.

Zue, V., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America, 66,* 1039–1050.