The influence of meaning on the perception of speech sounds

Nina Kazanina*†, Colin Phillips‡, and William Idsardi‡

*Department of Linguistics, University of Ottawa, 401 Arts Hall, 70 Laurier Avenue, Ottawa, ON, Canada K1N 6N5; and ‡Department of Linguistics and Neuroscience and Cognitive Science Program, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742

Communicated by Morris Halle, Massachusetts Institute of Technology, Cambridge, MA, June 9, 2006 (received for review January 16, 2006)

As part of knowledge of language, an adult speaker possesses information on which sounds are used in the language and on the distribution of these sounds in a multidimensional acoustic space. However, a speaker must know not only the sound categories of his language but also the functional significance of these categories, in particular, which sound contrasts are relevant for storing words in memory and which sound contrasts are not. Using magnetoencephalographic brain recordings with speakers of Russian and Korean, we demonstrate that a speaker's perceptual space, as reflected in early auditory brain responses, is shaped not only by bottom-up analysis of the distribution of sounds in his language but also by more abstract analysis of the functional significance of those sounds.

auditory cortex | native phonology | magnetoencephalography

Much research in speech perception has explored cross-language perceptual differences among speakers who have been exposed to different sets of sounds in their respective native languages. This body of work has found that effects of experience with different sound distributions are observed early in development (1–3) and are evident in early automatic brain responses in adults (e.g., ref. 4). In contrast, in this study we investigate how perceptual space is influenced by higher-level factors that are relevant for the encoding of words in long-term memory while holding constant the acoustic distribution of the sounds.

Recently, a number of proposals have suggested that the properties of a speaker's perceptual space can be derived from the distribution of sounds in acoustic space. According to such accounts, the learner discovers sound categories in the language input by identifying statistical peaks in the distribution of sounds in acoustic space. Recent evidence suggests that infants may indeed be able to carry out such distributional analyses (5, 6). However, such distributional analyses are of less use in determining how these sounds are linked to phonemes, the abstract sound-sized units that are used to encode words in memory. This is because there is not a one-to-one mapping between phoneme categories, the units used to store words, and the speech sound categories, sometimes known as phones, that are used to realize phonemes (7, 8). There are different possible mappings between phonemes and speech sounds, and therefore sets of sound categories with similar acoustic distributions may map onto different sets of phonemes across languages. A pair of sound categories in a language may be straightforwardly represented as a pair of different phonemes for purposes of word storage. Following standard notation, phonemes are represented by using slashes, and speech sounds/phones are represented by using square brackets, e.g., phoneme /p/ vs. speech sound [p].

For example, for an adult English speaker the first sound in words like *pin* or *pat* and the second sound in words like *spin* or *spam* correspond to the same phoneme /p/, and they are encoded identically in word storage. Yet word-initial /p/ and the /p/ that follows the sound "s" are systematically different in English speakers' speech production. Word-initial /p/ is aspirated (pronounced with a small burst of air after the consonant), but when /p/ follows /s/ it is unaspirated. In Thai, on the other hand, the aspirated and unaspirated sounds correspond to different phonemes, and there are many pairs of words in Thai that differ only in the presence or absence of aspiration on a "p" sound. So, the phones [p] and [p^h] exist in both Thai and English, but only in Thai do they correspond to distinct phonemes. In English, the two types of "p" are reliably differentiated in speech production yet correspond to a single memorized phoneme category. In standard terminology such sound pairs are known as allophonic categories or allophones of a single phoneme.

The focus of the current study is on a similar cross-language contrast, involving the mapping from the speech sound categories [d] and [t] onto phonemes. The aim is to test whether early stages of speech sound processing are governed by purely acoustic properties of the sound or whether they are affected by the functional role of the sound in word representations, i.e., by its phonemic status. Russian speakers systematically distinguish the sounds [d] and [t] in their speech production. These sounds show a bimodal distribution along the dimension of voice-onset time (VOT) (see *Materials and Methods*) and they are also used to encode meaning contrasts, as shown by minimally different word pairs like ***d**om* "house" vs. ***t**om* "volume" or *so**d**a* "baking soda" vs. *so**t**a* "cell." Thus, in Russian the sounds map onto distinct phoneme categories (Fig. 1). Korean speakers produce a very similar pair of sounds [d] and [t] in their speech with a similar bimodal acoustic distribution along the VOT dimension, but they never use these sounds to contrast word meanings. Korean [d] and [t] map onto a single phoneme, which we write here as /T/, and appear in complementary environments. Within words /T/ is realized as [d] between voiced sounds, as in *pa**d**a* "ocean" or *ka**d**ik* "full," and /T/ is realized as [t] elsewhere, e.g., in word-initial position in *tarimi* "iron" or *tarakbaŋ* "attic." Consequently, there are no pairs of words in Korean that differ only in the [d]/[t] sound (9–13).

The question of whether higher-level phoneme categories are derivable from the distribution of sounds in acoustic space is important in light of reports that higher-level categories may affect a speaker's perceptual space. Several behavioral studies have shown that discrimination of an allophonic contrast is poorer than discrimination of a phonemic contrast of an equivalent acoustic size (14, 15). This finding suggests that a speaker's perception of a sound contrast is affected by the status of that contrast at the level of word meanings. However, behavioral responses may reflect late, conscious processes and be affected by other systems such as orthography, and thus they may mask speakers' ability to categorize sounds based on lower-level acoustic distributions. Electrophysiological responses may help to resolve this issue because they can be collected continuously from the onset of the sound, and early response components such as MMN (mismatch negativity) and its magnetic counterpart MMNm have been shown to reflect preattentive processes (16).

Conflict of interest statement: No conflicts declared.

Abbreviations: MEG, magnetoencephalograph/magnetoencephalography; MMN, mismatch negativity; MMNm, magnetic MMN; VOT, voice-onset time.

†To whom correspondence should be addressed. E-mail: ninaka@uottawa.ca.

© 2006 by The National Academy of Sciences of the USA

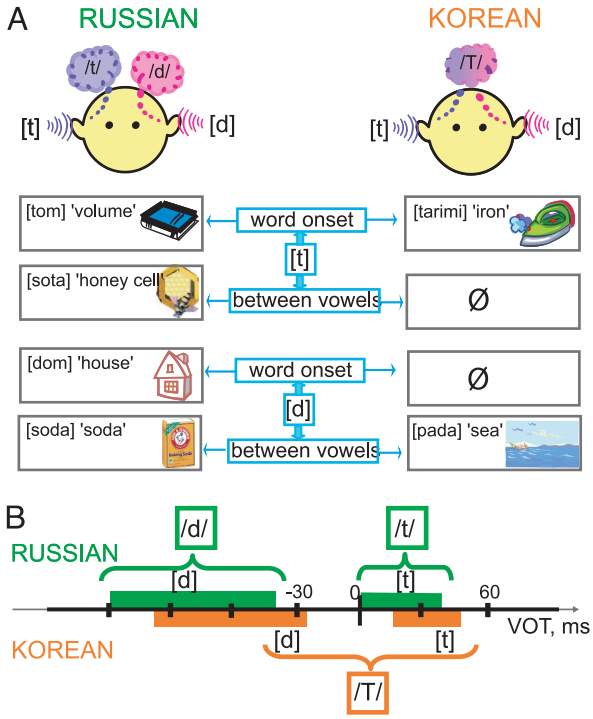

Fig. 1. The mental representation (*A*) and articulatory production (*B*) of the sounds [t] and [d] in Russian and Korean. (*A*) In Russian, syllable-initial [t] and [d] correspond to the distinct phonemes /t/ and /d/ and thus give rise to minimal pairs of words that contrast only in the [d]/[t] sound. In Korean, syllable-initial [t] and [d] are realizations of the same phoneme, here written as /T/. No Korean word pairs show a minimal contrast in the sounds [d]/[t]. No Korean words have [t] in an intervocalic position or [d] in word-initial position. (*B*) Approximate VOT values for intervocalic [d] and word-initial [t] in speakers of Korean (11, 13, 17) and Russian. Although the absolute VOT values differ between the two languages, production data show a clear bimodal distribution of tokens of [t] and [d] along the VOT continuum in both languages. VOT values shown for Russian [d] are based on intervocalic contexts, to allow for closer comparison with Korean values. VOT values for Russian word-initial [d] are similar.

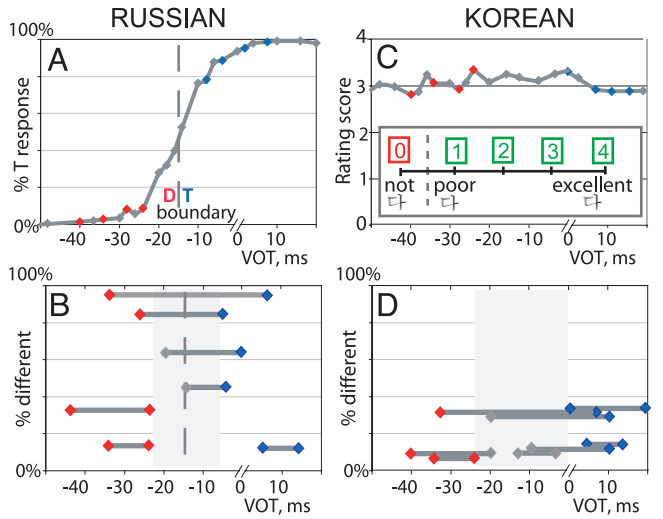

Fig. 2. Results from behavioral tests for Russian speakers (n = 13, mean age 25.6) and Korean speakers (n = 13, mean age 28.8). In each language, 0-ms VOT in the figures represents the "basic" token for that language (see *Materials and Methods*). (*A*) Russian speakers showed a classic identification function for syllables along the /da/-/ta/ VOT continuum with a crossover point at around −16-ms VOT. (*B*) Russian speakers more successfully discriminated syllable pairs when the members of the pair fell on opposite sides of the category boundary, relative to equidistant pairs of syllables that fell on the same side of the category boundary. (*C*) In contrast, Korean speakers showed no evidence of categorical perception in either task. In the naturalness rating task, they rated all syllables along the VOT continuum as equally natural instances of 다 /Ta/, for contextually natural positive VOTs and contextually unnatural negative VOTs alike. (*D*) Discrimination accuracy among Korean speakers was generally low. Unlike Russian, discrimination accuracy was better predicted by the acoustic distance between the sounds in each pair, rather than by the categorical status of the two sounds. In *A* and *C*, the four tokens indicated by blue and red represent the [ta] and [da] sounds used for the MEG experiment in each language. In *B* and *D*, the shaded area represents the gap between the standard and deviant categories in the MEG experiment.

We used whole-head magnetoencephalographic (MEG) brain recordings to measure the detailed time course of brain activity in speakers of Russian and Korean while they listened passively to the syllables [da] and [ta] in an oddball paradigm. In most previous studies the presence or the latency of an MMN response signaled the speaker's ability to discriminate an acoustic–phonetic contrast between the standard and the deviant tokens (4, 18–21). In contrast, by using a paradigm in which multiple nonorthogonally varying tokens from each category are presented (22), the presence of an MMNm in our study can serve as a measure of grouping of different acoustic tokens into phoneme or allophone categories. If the comparison of [d] and [t] categories elicits an MMNm response in Russian but not in Korean speakers, then we may conclude that category-based grouping of speech sounds can be performed based on phoneme but not on allophone categories, despite the fact that the distribution of the categories is bimodal in either language. Thus, an adequate model of the speaker's perceptual space would need to consider, in addition to the acoustic distributional properties of sounds, the functional status of sound contrasts for the purposes of encoding word-meanings.

Results

Both language groups participated in a pair of behavioral tests, followed by an MEG recording session on a later day. Participants heard all instructions in their native language. In the behavioral testing, the Russian group performed an identification task, whereas the Korean group performed a naturalness rating test (Fig. 2*A*). An identification task was not possible for Korean speakers because their language does not provide distinct orthographic labels for [d] and [t] sounds. In addition, both language groups performed an AX discrimination task (Fig. 2*B*). The AX paradigm was chosen because it provided the best opportunity for Korean speakers, who lack a phonemic representation of the voicing contrast, to demonstrate discrimination sensitivity based on purely acoustic factors.

All MEG sessions began with a screening study that verified the presence of a clear auditory N100m response in each participant. In the Speech condition Russian and Korean participants were exposed to multiple nonorthogonally varying tokens of the syllables [da] and [ta] presented in an oddball paradigm (see *Materials and Methods* and Fig. 3*B*). This design provides a measure of category formation: A mismatch response can only be elicited if the different exemplars of each category are treated as equivalent at some level. In the Tone condition, participants listened to contrasting sinusoidal tones (Fig. 3*C*). This condition was included to control for the possibility that any cross-language differences in the results of the speech condition might reflect between-group differences in basic auditory evoked responses.

Behavioral. In the identification task, Russian speakers showed categorical perception of [d] and [t] categories along the VOT continuum (Fig. 2*A*), and in the discrimination task, equidistant pairs of syllables were discriminated significantly more accu-

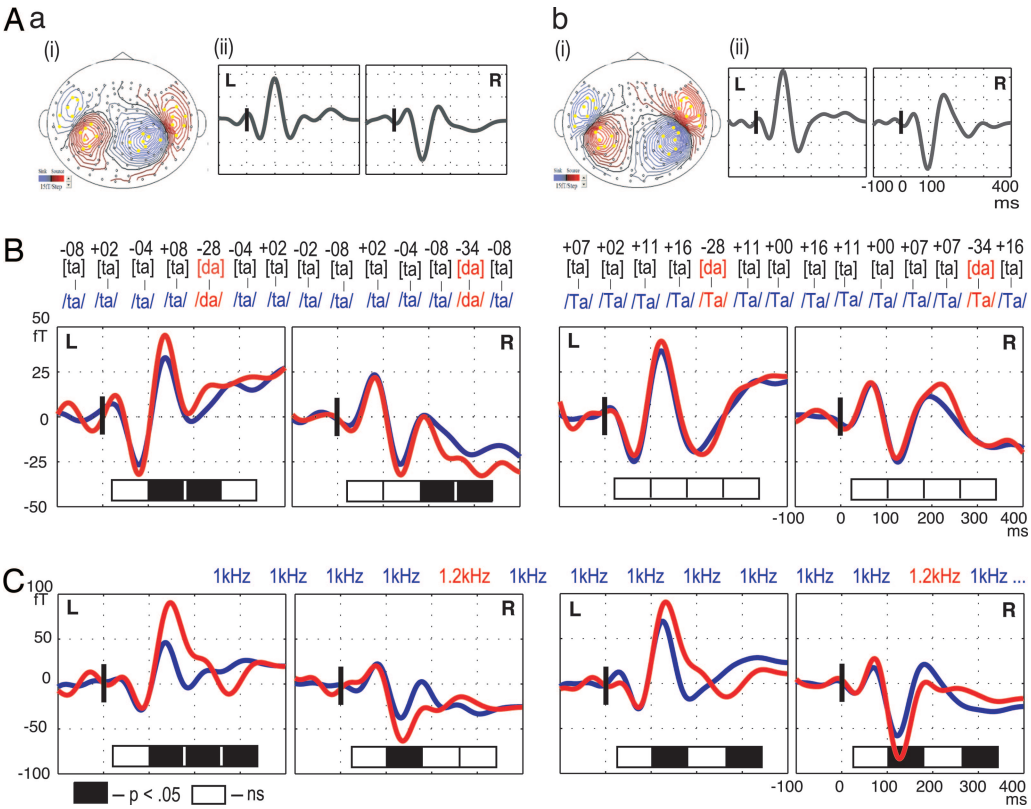

Fig. 3. Tone pretest, Speech condition, and Tone condition. (*A*) Tone pretest. (*Ai*) Averaged evoked response to 1-kHz tone in Russian (*a*) and Korean (*b*) language groups, showing dipolar scalp distribution of N100m response in each hemisphere for a representative participant in each group. Yellow dots indicate the seven MEG channels with the largest N100m amplitude at each field maximum. For each participant, these channels were used for the analysis of the Tone and Speech conditions. (*Aii*) Grand average N100m response from seven posterior channels per hemisphere ($n = 13$ in each group). (*B*) Speech condition. Each language group listened to sequences of syllables drawn from similar [da] and [ta] categories, each consisting of multiple tokens with different VOT values (upper row). The acoustic and phonetic encoding of the syllable sequences was comparable across languages, but Korean lacks the contrast at the phoneme level because [d] and [t] are allophones of the single phoneme category /T/. The waveforms represent the grand average brain response from seven left-posterior and right-posterior channels across two blocks. Black rectangles in each graph indicate the locus of significant differences, as indicated by a one-way ANOVA performed in four 80-ms-long windows (from 20 to 340 ms) within individual scalp quadrants. Russian speakers showed a significant divergence between responses to standard (blue) and deviant (red) categories, whereas Korean speakers did not. Because of the use of multiple nonorthogonally varying tokens of each category, a many-to-one ratio among stimuli was absent at the acoustic level for both language groups and available at the phonemic level only for Russian speakers. Therefore, the presence of an MMNm in Russian but not in Korean speakers provides evidence that auditory cortex can group sounds based on phonemic but not allophonic categories. (*C*) Tone condition. For the same MEG channels in the same two groups of speakers, grand averaged brain responses to a standard 1-kHz (blue) and a deviant 1.2-kHz (red) tone are shown. A typical N100m response followed by an MMNm to the deviant tones was found in both language groups.

rately if the syllables fell on opposite sides of the category boundary (Fig. 2*B*). Korean speakers showed no evidence of categorical perception of [d] and [t] categories. In the naturalness rating task, all syllables received high ratings as instances of the category /Ta/, despite the fact that the negative VOT values were contextually inappropriate (Fig. 2*C*). Korean speakers also showed low accuracy in the discrimination task, even for pairs containing VOT values that speakers clearly differentiated in their productions, e.g., [da] with VOT −34 ms vs. [ta] with VOT +8 ms (Fig. 2*D*).

MEG. Statistical analyses of the Speech and Tone conditions were performed separately for each language group. ANOVAs with the within-subject factors category (standard vs. deviant), hemisphere (left vs. right), and anteriority (anterior vs. posterior) were calculated based on mean magnetic field strengths in four 80-ms time intervals: 20–100, 100–180, 180–260, and 260–340 ms. In addition, separate analyses of the effect of the category factor were carried out for anterior and posterior regions within each hemisphere. Successful classification of the sounds from the two sound categories in each condition should be reflected in a mismatch response (MMNm). Because of the dipolar distribution of MEG responses recorded by using axial gradiometers, the mismatch response should be reflected in an effect of category that shows opposite polarity at anterior and posterior field maxima. Because of the reversal of magnetic field distributions across hemispheres in the auditory evoked response (Fig. 3*A*), the polarity of the mismatch response should also reverse across hemispheres. We report here all significant main effects and interactions that involve the category factor. The text and Fig. 3 focus on effects of the category factor at groups of MEG channels that represent individual field maxima, but evidence for a mismatch response is also provided by the three-way interaction category × hemisphere × anteriority in the overall ANOVA.

Speech Condition. In the Speech condition, participants heard sounds drawn from the [d] and [t] categories in a many-to-one ratio. Each speech category was represented by multiple exemplars that varied along the same VOT dimension that distinguished the two categories. Unlike the Tone condition reported below, the Speech condition showed a clear contrast between the two language groups (Fig. 3*B*).

Russian. In the Russian group, the deviant category elicited an early left-lateralized mismatch response, starting at ≈100 ms. Analyses of MMNm at individual scalp quadrants indicated an earlier onset in the left than in the right hemisphere. At the left posterior quadrant, the effect of category was present in the 100- to 180-ms interval [$F(1, 12) = 8.4$, $P < 0.05$] and continued into the 180- to 260-ms interval [$F(1, 12) = 5.7$, $P < 0.05$]. At the right posterior quadrant, the effect of category was significant only in the 180- to 260-ms and 260- to 340-ms intervals [$F(1, 12) = 13.4$, $P < 0.01$ and $F(1, 12) = 9.5$, $P < 0.001$, respectively]. The mismatch response was also reflected in the overall ANOVA as a significant three-way interaction at all intervals from 100 to 340 ms [100–180 ms: $F(1, 12) = 10.1$, $P < 0.01$; 180–260 ms: $F(1, 12) = 11.5$, $P < 0.01$; 260–340 ms: $F(1, 12) = 5.6$, $P < 0.05$]. There was also a two-way interaction of category and anteriority in the 340- to 420-ms interval [$F(1, 12) = 6.2$, $P < 0.05$].

Korean. In the Korean group, there were no significant main effects or interactions involving the category factor (all P values > 0.1). Planned comparisons of the standard and deviant categories in each scalp quadrant revealed no significant effects of category at any time interval.

Tone Condition. In the Tone condition, both language groups exhibited a highly reliable mismatch response to the deviant 1.2-kHz tone relative to the standard 1-kHz tone in the 100- to 180-ms time interval, and this effect persisted to subsequent time intervals in both language groups in at least one hemisphere (Fig. 3*C*). These results suggest that neither group shows inherently stronger auditory evoked responses.

Discussion

The aim of this study was to assess the relative contribution to perceptual categorization of the acoustic distribution of sounds in a speaker's linguistic experience on the one hand and the functional significance of those sounds for encoding word meanings on the other hand. Although Russian and Korean speakers both exhibit a bimodal acoustic distribution of the sounds [d] and [t] in their speech production, only in Russian are these sounds used contrastively for encoding word meanings. In an oddball paradigm, we recorded MEG brain responses to multiple instances of sounds drawn from the [d] and [t] categories and found that Russian speakers showed evidence of rapid separation of these sounds into two categories, as signaled by a mismatch response following the N100m response. In contrast, the Korean group showed no immediate sensitivity to the difference between the two categories. This was despite the fact that the clusters of [d] and [t] tokens presented to Korean speakers had a greater acoustic between-category separation along the VOT continuum (see *Materials and Methods*), which might have made it easier for the Korean speakers to separate the two categories of sounds. The cross-language contrast is particularly notable given that the same Korean speakers who participated in the MEG experiment systematically differentiate the two sounds in their own speech, producing [d] in intervocalic contexts and [t] elsewhere.

Speech sounds are characterized by substantial variability. When processing speech, a speaker must disregard some variation in the input and preserve other distinctions as highly relevant for purposes of word recognition. The question then arises of the relation between the code used for storing words (i.e., the phonemes of the language), the statistical distribution of the sounds of the language in acoustic space, and the preattentive perceptual abilities of the speaker. A popular approach is to assume that a speaker's perceptual categories are formed as a result of a bottom-up analysis of the acoustic properties of sounds in the input (e.g., refs. 6, 23, and 24). Under this approach, the speaker's perceptual map is largely shaped by the distribution of sounds in acoustic space, and phonemes mirror acoustic category prototypes corresponding to the peaks of Gaussian distributions, where the highest concentration of tokens is observed. Areas around each peak correspond to less-frequent and less-prototypical instances of the same category. Sounds that belong to different clusters in acoustic space are predicted to be perceptually distinct, whereas tokens that are part of the same distributional cluster yield the same perceptual object. Previous electrophysiological studies on speech perception did not employ allophonic contrasts and are compatible with this bottom-up approach. Early automatic brain responses are affected by whether a sound corresponds to a category prototype in the speaker's native language (4) or whether a sound contrast corresponds to a phoneme contrast in the language (18, 20, 25). These earlier results could be explained by assuming that discrimination of sounds that belong to the same cluster in acoustic space is poor, whereas discrimination of sounds from different acoustic clusters is enhanced.

A unique aspect of this study is that, to further determine the source of perceptual categories, we considered cases in which there is no one-to-one correspondence between maxima in acoustic space and the higher-order phonemic categories used to store words. Across languages, it is rather common for a phoneme to be realized by two or more allophonic variants, where each allophonic category corresponds to a different cluster in acoustic space. In the current case, the Korean sounds [d] and [t] realize the same phoneme category yet form a bimodal distribution in the acoustic space. If the speaker's perceptual map were simply a direct reflection of acoustic space, Korean speakers should be able to form perceptual categories based on groups of [d] and [t] sounds just as well as Russian speakers. Indeed, they must do so during language learning, to properly encode the pronunciation rules of the language. Yet the comparison of Russian and Korean adult speakers in both behavioral tasks and electrophysiological measures does not bear out this prediction. Russian speakers exhibit categorical perception of the [d]/[t] contrast in every task. Korean speakers, in contrast, show great difficulty in consciously discriminating the two sound categories, and their brain responses in the oddball paradigm suggest that allophonic categories fail to serve as a basis for categorization and grouping of sounds. Results of the MMNm study in the two language groups suggest that auditory cortical systems may disregard acoustic distinctions that are statistically reliable in the input to the speaker, unless those distinctions are relevant for the encoding of word meanings. Thus, the speakers' perceptual space is shaped not only by bottom-up analysis of the distribution of sounds in a language but also by more abstract analysis of the functional significance of those sounds.

The lack of a mismatch response to the [d]/[t] contrast in Korean speakers points specifically to a limited role for distributionally defined categories. Our modified oddball paradigm introduced substantial within-category variation that was nonorthogonal to the variation between the standard and deviant categories (22), and therefore the elicitation of a mismatch response was predicated upon correct separation of the sounds into two distinct groups. Our MEG results indicate that auditory cortex supports grouping based on phoneme categories (Russian) but not on allophone categories (Korean). The fact that a mismatch response was elicited only in the Russian group, for which a physical separation between the standard and the deviant categories was smaller than in the Korean group (see *Materials and Methods*), suggests that given the current nonorthogonal design the mismatch in the Russian group cannot be explained solely by acoustic differences between the standards and the deviants. Previous studies have shown that a mismatch response may be elicited by a subphonemic acoustic contrast. However, because these studies were

based on designs in which standard and deviant sounds correspond to a single fixed stimulus (e.g., refs. 18 and 20), they speak to the issue of fine-grained acoustic discrimination rather than to the issue of the formation of perceptual categories. The results of our behavioral tests are consistent with previous behavioral reports of poor discrimination of allophonic contrasts (14, 15). Together with the current MEG results, which cannot be attributed to late influences of conscious perceptual categories such as orthographic conventions of the language, our findings lead to the conclusion that the phonemic code used to store words in memory exerts an immediate effect on perceptual categorization, and thus they strongly support the hypothesis that representations that are immediately computed from speech are phonemic in nature (26–28).

Materials and Methods

Participants. Participants were healthy adult native speakers of Russian or Korean with no previous history of hearing problems or language disorders, all of whom were born in and had lived at least 18 years in Russia or Korea and had spent no more than 4 years outside their home country. The mean length of stay in the United States was 2.0 years for the Russian group and 1.7 years for the Korean group. The mean age was 25.6 years for the Russian group ($n = 13$, 7 males, range 18–33 years) and 28.8 years for the Korean group ($n = 13$, 5 males, range 26–33 years). An additional three Russian and four Korean participants were excluded based on the lack of a strong bilateral N100m response elicited by a 1-kHz pure tone in a pretest. All participants were strongly right-handed, gave informed consent, and were paid $10 per hour for their participation.

Stimuli. In each language, stimuli were chosen from a single [da]-[ta] continuum that varied in VOT. VOT is a measure of the interval between the consonant release and the onset of vocal fold vibration (29). VOT values are positive if the consonant release precedes the onset of vocal fold vibration and negative otherwise. In intervocalic position, the VOT is set to the duration of the fully voiced closure immediately before stop consonant release (cf. ref. 18). For each language, the continuum was constructed from a recording of a single basic /ta/ syllable, spoken by a male speaker of each language, ensuring that the vowel was a natural instance of /a/ in either language. The negative range of the VOT continuum was generated from the basic token by adding increments of a periodic voicing lead. The voicing lead in both languages was drawn from a natural recording of a Korean speaker, thereby biasing the materials against the hypothesized perceptual advantage for Russian speakers. The positive range of the VOT continuum was generated by splicing in intervals of low-amplitude aspiration between the consonant release and the onset of periodic voicing. The continuum covered values from −50 to +20 ms VOT. The stimuli were digitized at a sampling rate of 44.1 kHz. In both the behavioral and MEG tests, the stimuli were delivered binaurally via insert earphones (Etymotic ER-3A; Etymotic Research).

Behavioral Tests. Each language group participated in two behavioral tests. The Russian group was administered an identification and a discrimination task. In the identification task, each participant heard individual syllables from the [da]-[ta] continuum and matched them to choices presented in Cyrillic script as *ДА* "DA" or *ТА* "TA." In the discrimination task, Russian participants gave same/different judgments about pairs of syllables drawn from the VOT continuum with 10- to 40-ms VOT distance between the tokens in a pair. The Korean group performed the same discrimination task as the Russian group but could not be administered the identification task used with Russian speakers, because the consonants [d] and [t] are not distinguished phonemically in Korean and are written using the same Hangul symbol ㄷ. The identification task was therefore replaced by a rating task, in which Korean speakers rated the naturalness of tokens from the continuum as a representative of the syllable 다 = 다 "Ta" on a 0-to-4 scale (0, not 다; 1, poor 다; 2, mediocre 다; 3, good 다; 4, excellent 다). To ensure that speakers would use the full rating scale, the task included other types of voiceless alveolar stops found in Korean, i.e., glottalized and aspirated stops, in addition to the plain stop ㄷ that undergoes intervocalic voicing.

MEG Recordings. Magnetic fields were recorded by using a whole-head MEG device with 160 axial gradiometers (Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 1 kHz. The recordings were divided into intervals of 500-ms duration, starting at 100 ms relative to the stimulus onset. Epochs that contained eye blinks and movement artifacts or that exceeded a threshold of 2 pT were excluded. The data were averaged, baseline-corrected by using the prestimulus interval, and filtered by using a 0.5- to 30-Hz band-pass filter. Data figures reflect the application of an additional 15-Hz low-pass filter, which was used for illustrative purposes only.

Each recording session began with a screening run, in which each participant's response to 200 repetitions of a 50-ms, 1-kHz sinusoidal tone was recorded. Only those participants who showed a strong bilateral N100m response in each hemisphere (see Fig. 3*A*) were included in further testing. The scalp distributions of averaged evoked N100m fields were consistent with the typical source of the N100m in supratemporal auditory cortex (cf. ref. 30). The screening run verified that the amplitude of auditory evoked responses was closely matched across the two language groups and provided an unbiased measure of the MEG channels to be included in analyses of the oddball conditions. For each hemisphere and each participant, 14 MEG channels were selected (i.e., 28 per participant), 7 each for the ingoing and outgoing magnetic fields, based on the strongest average field at the N100m peak.

The two blocks of the Speech condition and one block of the Tone condition were run using an oddball paradigm with a many-to-one ratio of standard to deviant stimuli (16). Each block consisted of 700 instances of the standard sound interspersed with 100 deviant sounds. In the Tone condition, the standard and deviant sounds were 1- and 1.2-kHz sinusoidal tones, respectively. The averaged response to the standard stimulus was then compared with the averaged response to the deviant stimulus. In the Speech condition, the category that was the standard in the first block (e.g., /ta/) was the deviant in the second block, with the order of blocks counterbalanced across participants. The standard and deviant categories were each represented by four tokens of varying VOT, presented randomly and with an equal probability. In Russian, the categories /da/ and /ta/ were equidistant from the category boundary located at −16 ms VOT (/da/-tokens: −40, −34, −28, and −24 ms; /ta/-tokens: −08, −04, +02, and +08 ms VOT). In Korean, a more conservative approach was followed, because of the absence of a psychophysically determined category boundary. All members of the [ta] category fell into the positive range of the VOT continuum ([da]-tokens: −40, −34, −28, and −24 ms; [ta]-tokens: 00, +07, +11, and +16 ms VOT). In either language, the between-category and the within-category variation lay along the same VOT dimension and thus were nonorthogonal. This design ensures that there is no many-to-one ratio among standards and deviants at an acoustic level and thus provides a measure of category identification (22). In Russian the 16-ms VOT variation within

each category was equal to the minimum VOT difference between the categories, whereas in Korean the within-category variation was smaller than the between-category distance (16 vs. 24 ms VOT, respectively). In purely acoustic terms, therefore, the larger between-category distance should have made it easier for Korean participants than for Russian participants to recognize the [da] and [ta] sounds as being drawn from a bimodal distribution. For each language group, we compared the averaged brain responses to the standard and deviant categories, combining data from both blocks of the Speech condition, such that the standard and deviant averages were based on exactly the same speech sounds.

We thank David Poeppel and Grace Yeni-Komshian for valuable comments; Jeff Walker and Sang Kim for assistance in running the experiments; and Soo-Min Hong, Slava Merkine, and Matt Reames for help in stimulus preparation and recruiting participants. We are extremely grateful to Kanazawa Institute of Technology, Japan, for equipment support. This work was supported by James S. McDonnell Foundation Grant 99-31T, Human Frontier Science Program Grant RGY-0134, and National Science Foundation Grant BCS-0196004 (all to C.P.).

1. Werker, J. F. & Tees, R. C. (1984) *J. Acoust. Soc. Am.* **75,** 1866–1878.
2. Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992) *Science* **255,** 606–608.
3. Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K. & Naatanen, R. (1998) *Nat. Neurosci.* **1,** 351–353.
4. Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., *et al.* (1997) *Nature* **385,** 432–434.
5. Kuhl, P. K. (2004) *Nat. Rev. Neurosci.* **5,** 831–843.
6. Maye, J., Werker, J. & Gerken, L. (2002) *Cognition* **82,** B101–B111.
7. de Saussure, F. (1916) *Cours de Linguistique Générale* (Payot, Paris).
8. Baudouin de Courtenay, J. N. (1972) *A Baudouin de Courtenay Anthology; The Beginnings of Structural Linguistics*, ed. and trans. Stankiewicz, E. (Indiana Univ. Press, Bloomington).
9. Martin, S. (1951) *Language* **27,** 519–533.
10. Kim-Renaud, Y.-K. (1974) Ph.D. dissertation (Univ. of Hawai'i, Manoa).
11. Silva, D. (1992) Ph.D. dissertation (Cornell Univ., Ithaca, NY).
12. Jun, S.-A. (1998) *Phonology* **15,** 189–226.
13. Cho, T. & Keating, P. (2001) *J. Phonetics* **28,** 155–190.
14. Pegg, J. E. & Werker, J. F. (1997) *J. Acoust. Soc. Am.* **102,** 3742–3753.
15. Whalen, D. H., Best, C. T. & Irwin, J. (1997) *J. Phonet.* **25,** 501–528.
16. Näätänen, R., Gaillard, A. W. K. & Mäntysalo, S. (1978) *Acta Psychol.* **42,** 313–329.
17. Yun, G. & Jackson, S. R. (2004) in *MITWPL 46: The Proceedings of the Workshop on Altaic in Formal Linguistics*, eds. Csirmaz, A., Walter, M.-A. & Lee, Y. (MIT Press, Cambridge, MA), pp. 195–207.
18. Keating, P. A., Mikos, M. J. & Ganong, W. F., III (1981) *J. Acoust. Soc. Am.* **70,** 1261–1271.
19. Sharma, A. & Dorman, M. (2000) *J. Acoust. Soc. Am.* **107,** 2697–2703.
20. Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., *et al.* (1999) *Brain Res. Cognit. Brain Res.* **7,** 357–369.
21. Eulitz, C. & Lahiri, A. (2004) *J. Cognit. Neurosci.* **16,** 577–583.
22. Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M. & Roberts, T. (2000) *J. Cognit. Neurosci.* **12,** 1038–1055.
23. Kohonen, T. & Hari, R. (1999) *Trends Neurosci.* **22,** 135–139.
24. Kuhl, P. K. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 11850–11857.
25. Dehaene-Lambertz, G. (1997) *NeuroReport* **8,** 919–924.
26. Whalen, D. H. & Liberman, A. M. (1987) *Science* **23,** 169–171.
27. Dehaene-Lambertz, G. & Gliga, T. (2004) *J. Cognit. Neurosci.* **16,** 1375–1387.
28. Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V., Lai, S., Mencl, W. E., Fulbright, R. K., Constable, R. T. & Liberman, A. M. (2006) *J. Acoust. Soc. Am.* **119,** 575–581.
29. Lisker, L. & Abramson, A. S. (1964) *Word* **20,** 384–422.
30. Alho, K. (1995) *Ear Hear.* **16,** 38–50.