

Identification of Phonemes: Differences between Phoneme Classes and the Effect of Class Size

Anita Wagner^a Mirjam Ernestus^{a,b}

^aMax Planck Institute for Psycholinguistics, and ^bRadboud University Nijmegen,
Nijmegen, The Netherlands

Abstract

This study reports general and language-specific patterns in phoneme identification. In a series of phoneme monitoring experiments, Castilian Spanish, Catalan, Dutch, English, and Polish listeners identified vowel, fricative, and stop consonant targets that are phonemic in all these languages, embedded in nonsense words. Fricatives were generally identified more slowly than vowels, while the speed of identification for stop consonants was highly dependent on the onset of the measurements. Moreover, listeners' response latencies and accuracy in detecting a phoneme correlated with the number of categories within that phoneme's class in the listener's native phoneme repertoire: more native categories slowed listeners down and decreased their accuracy. We excluded the possibility that this effect stems from differences in the frequencies of occurrence of the phonemes in the different languages. Rather, the effect of the number of categories can be explained by general properties of the perception system, which cause language-specific patterns in speech processing.

Copyright © 2008 S. Karger AG, Basel

Introduction

Listeners are able to focus on individual speech sounds and identify them in an effortless and largely accurate manner. Here we investigate whether identification of speech sounds varies among sound classes and among listener groups with different sets of contrastive speech sounds. We compare the identification of speech sounds between vowels, fricatives, and stop consonants and across listeners with a vowel- or fricative-rich repertoire versus listeners with fewer categories in these speech sound classes.

Models of speech perception vary in the role they ascribe to individual speech sounds, and whether they incorporate a level of prelexical phonemic processing [e.g., McClelland and Elman, 1986; Norris et al., 2000; Johnson, 2004]. While the instant activation of phonemes in speech processing is controversial, daily observations, such

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2008 S. Karger AG, Basel
0031–8388/08/0652–0106
\$23.50/0
Accessible online at:
www.karger.com/journals/pho

Anita Wagner
Max Planck Institute for Psycholinguistics
PO Box 310
NL-6500 AH, Nijmegen (The Netherlands)
Tel. +31 24 352 1585, Fax +31 24 352 1213
E-Mail Anita.Wagner@mpi.nl

as the occurrence of spoonerisms and puns in languages, and phonemically based orthographic systems, show listeners' effortless ability to focus on individual speech sounds. Furthermore, several studies have demonstrated that listeners' perception adjusts rapidly to speaker-specific phoneme realisations [Norris et al., 2003; Eisner and McQueen, 2005], and such adjustments spread to other instances of these phonemes in new words [McQueen et al., 2006]. Moreover, brain imaging studies have shown the existence of neuronal traces of phoneme representations [Näätänen et al., 1997]. Also, reports on brain-damaged patients show that listeners may have a normal ability to recognize individual speech sounds even though their lexical representations are disrupted [e.g., Martin et al., 1999].

Commonly, speech sounds are divided into two main classes: vowels and consonants. These two groups are the alternating building blocks of words. They differ in their phonological function, with vowels forming the centres and consonants forming the margins of syllables. The different phonological functions of vowels and consonants are reflected in different contributions of these speech sound classes to word recognition: vowels appear to restrict lexical selection less than consonants [Cutler et al., 2000; Bonatti et al., 2005]. Cutler et al. [2000], for instance, showed that speakers tend to change vowels rather than consonants when they are asked to turn pseudo-words into existing words. Further indications for differences in processing between vowels and consonants come from aphasic patients: patients may be hampered in the production of only one of these classes, suggesting that vowels and consonants are processed by distinct neural mechanisms [Caramazza et al., 2000; but see Sharp et al., 2005, for a different view in perception].

In acoustic terms, stop consonants are very different from vowels, and this acoustic difference forms the basis of the explanation for categorical perception of stop consonants versus continuous perception of vowels. In a series of identification and discrimination experiments, Liberman et al. [1957] observed that listeners perceive stop consonants categorically (i.e., do not distinguish between different realisations of the same phoneme), whereas differences in the precise quality within a vowel category are perceived easily (more continuous discrimination). The perception of intraphonemic acoustic variation in stop consonants is less in correspondence with the actual fine-grained variation in the acoustic signal than the perception of subtle acoustic differences within a vowel category.

Pisoni and Tash [1974] suggested that vowels and consonants differ in the way they are encoded in auditory and phonetic memory. As argued by Pisoni [1973], two modes of memory play a role in phoneme discrimination and identification: auditory memory, where detailed perceptual traces are stored but decay fast, and phonetic memory, where the acoustic signal is assigned to phonemic categories. Stop consonants, because of their shorter and more abrupt acoustic properties, leave traces in auditory short-term memory that decay faster compared to the traces of longer and continuous acoustic events like vowels. As a consequence, the traces of vowels are longer available for retrieval, and they allow detailed and more continuous discrimination. When discriminating stop consonants, listeners rely more on the information in phonetic memory, where the signal has been labeled and assigned to a phonemic category.

If the difference between categorical and continuous perception is due to the acoustic properties of the segment, the large group of consonants should also show within-group differences, as this group contains a heterogeneity of phonemes with many different acoustic properties. This is indeed the case. Healy and Repp [1982]

conducted identification, discrimination, and labelling experiments with vowels and fricatives, and found that, in contrast to stop consonants, both vowels and fricatives are not categorically perceived. The discrimination precision was even higher for fricatives than for vowels.

Two decades later, Mirman et al. [2004] investigated the processing of non-speech sounds. Listeners categorised non-speech materials, which contained either steady-state sounds resembling simplified vowels or fricatives, or sounds with transient properties similar to consonants like stop consonants, or both. It appeared that listeners cannot discriminate rapidly changing sounds belonging to the same category, while they can easily perceive subtle acoustic variation within the boundaries of a category for steady-state sounds. The authors conclude that this supports the hypothesis that vowels and fricatives are identified differently from stop consonants because of their acoustic properties. Rapidly changing sounds, such as stop consonants, tend to be discriminated according to their phonemic labels, while steady-state sounds, such as vowels and fricatives, tend to be discriminated in an acoustically more detailed manner.

Differences between vowels, fricatives, and stop consonants are also reflected in response latencies in phoneme monitoring experiments. Foss and Swinney [1973] reported slightly longer response times to fricatives than to stop consonants. Similarly, Savin and Bever [1970] found that listeners identify an initial phoneme in nonsense syllables faster if it is a stop consonant than if it is a fricative, while vowels are detected even more slowly. Rubin et al. [1976] observed similar differences between word-initial /b/ and /s/ and Morton and Long [1976] between word-initial plosives and non-plosives, which included fricatives, glides, and a nasal. Finally, Van Ooijen [1994] showed that the position of the phoneme in the word may play a role, at least if the stimulus is an existing word. She found that vowels were detected more slowly than stop consonants and fricatives, especially in word-final position.

Note that the studies summarised above all investigated phoneme recognition with native speakers of English. A study by Cutler and Otake [1994] is exceptional in this respect. It compared the identification of nasal consonants and vowels by Japanese and English listeners. English listeners detected vowels significantly more slowly and less accurately than nasals, independently of whether these sounds were presented in English or in Japanese words. Japanese listeners, on the other hand, did not recognise vowels more slowly than nasals. Cutler and Otake [1994] argued that Japanese listeners are not slower in identifying vowels than consonants because, in contrast to English listeners, they have only few vowels in their phoneme inventory with which a target vowel can be confused. Language-specific properties may thus obscure or induce seemingly general differences between phoneme classes, since listeners' perception is shaped by their experience with their native speech sound categories.

Also Costa et al. [1998] have reported that the number of phonemes in the native inventory plays a role in phoneme identification. The authors described a phoneme monitoring experiment with Dutch and Spanish participants. Listeners detected vowel or consonant targets in CVCVCVCVCV strings, in which the vowel or the consonant preceding the target was either constant over the stimulus or varied between syllables (e.g., for the target /p/ *ku su tu su pu* versus *ko se to si pu*). Dutch listeners, whose language has an approximately balanced vowel-to-consonant ratio, were delayed to the same extent by variation in the consonantal context for vowels as by variation in the vocalic context for consonant targets. In contrast, Spanish listeners, whose phoneme repertoire has four times as many consonants as vowels, showed a greater effect of

variation in the consonantal than in the vocalic context. Costa et al. [1998] explained this difference between Dutch and Spanish by arguing that listeners are aware of the influence that co-occurring phonemes have on the exact realisation of a phoneme. For consonants, this variation is smaller in Spanish than in Dutch, as Spanish has only five, instead of 16 vowels.

Combining the findings in these studies on the processing of speech sounds, we formulated two hypotheses. Both hypotheses may affect listeners' identification of speech sounds simultaneously. The first hypothesis states that speech sound classes require different recognition times. This hypothesis is based not only on the differences in acoustic properties between the sound classes but also on differences in phonological and lexical function. As mentioned above, vowels play a smaller role in lexical processing than consonants, and reaction times may therefore be longer for vowels.

The second hypothesis is that differences between the speech sound classes will be modulated by the number of categories within these classes in the listener's native phoneme repertoire. Listeners with a higher number of categories within a certain speech sound class will identify a target of that class more slowly than listeners whose native repertoire does not contain as many categories in that class. If this hypothesis is correct, the number of categories should be taken into account in order to ascertain general differences between vowels, stop consonants, and fricatives.

Importantly, the second hypothesis is based on general processes of categorisation, which are not restricted to auditory perception. When participants make decisions, like for instance about the identity of a colour or shape, their processing time is longer when they have more alternative choices [e.g., Hick, 1952; Theeuwes, 1992; Schweickert, 1993; Nosofsky, 1997]. In order to make clear that our second hypothesis is not specific for speech processing, we will use the term 'category' to refer to phonemes. Categories instantiate listeners' knowledge, which may be formulated in terms of phonemes, and which is established during speech development. Note that even though the effect of the number of categories would result in language-specific performance, it would affect listeners of all languages in the same way.

The question arises whether the relevant categories are indeed the phonemes. Many phonological and psycholinguistic models [e.g., McClelland and Elman, 1986; Norris et al., 2000] assign an important role to the phoneme, which is a theoretical construct. Listeners, however, can also distinguish between allophones of the same phoneme [e.g., between the palatal and the uvular fricative in German, see Lipski, 2006], and these allophones may therefore play an important role in speech processing as well. Hence, the number of relevant categories may be the number of phonemes or the number of distinguishable speech sounds. We decided to focus on phonemic categories in the current study. The most important reason is that there is not sufficient data to determine which sounds can be distinguished by which listeners.

Different from the studies mentioned above, our study examined five listener groups of different native backgrounds (in previous studies maximally two groups had been tested). If indeed phoneme classes differ in the speed and accuracy of identification due to their function and acoustic manifestation, the same differences should be found for all languages. However, as the listener groups differ in their number of categories for these phoneme classes, we nevertheless expect differences between the language groups, as a function of these numbers of categories.

Naturally, the languages of the listeners also differ in many other respects, in addition to their phoneme inventories, and these differences contribute to differences in speech

processing. Examples are the languages' stress patterns, syllable structures, and phonotactic restrictions. These language-specific characteristics might make it difficult to find clear general differences between phoneme classes and a role for the number of categories.

In order to investigate how listeners' perception is shaped by both general and language-specific factors, we have to make sure that all listeners can use their native listening strategies. One possibility is to present each listener group with natural materials produced by a native speaker of their own language. This, however, would introduce an additional source of variability, as all language groups would then be presented with different stimuli. Another possibility is to present all listeners with synthetic stimuli, which has frequently been done in cross-linguistic research [e.g., Bradlow, 1996; Iverson et al., 2003]. With synthetic stimuli, however, we run the risk of presenting listeners with impoverished stimuli. Previous findings show that listeners differ in their selection of, or attention to, acoustic cues, depending on their native language [Iverson et al., 2003; Wagner et al., 2006], and synthetically generated materials may fail to represent especially those cues relevant only for some groups of listeners.

We decided to take advantage of the assumption that listeners, when presented with a foreign language, assign the foreign sounds to their most similar native categories. We presented all listeners with the same naturally produced materials, consisting of segments which are phonemic in all languages to be tested. Variability between language groups was further reduced by choosing nonsense words as materials. In such a way, we restricted potential lexical effects and created conditions under which listeners focus more on the acoustic surface form of the materials.

An experimental paradigm that can reveal processes at the level of speech sounds by means of nonsense words is phoneme monitoring. In this paradigm, listeners are presented with lists of words, sentences, or nonsense words, and are asked to detect target phonemes. The measured reaction times and accuracy can give us insight into speech processing, including general and language-specific patterns in speech perception [for an overview see Connine and Titone, 1996]. Thus, with this paradigm, language-specific strategies can be revealed, and have previously been reported, for speech perception [Cutler and Otake, 1994; Costa et al., 1998].

Phoneme monitoring is a much-used paradigm that has contributed to the investigation of a wide range of questions, regarding both the prelexical and the lexical level of speech processing. Results obtained with phoneme monitoring have been replicated by means of other experimental paradigms, especially auditory lexical decision, such as the role of a word's frequency of occurrence [phoneme monitoring: Dupoux and Mehler, 1990; lexical decision: Luce, 1986] and phonological similarity effects [phoneme monitoring: Foss and Dowell, 1971; lexical decision: Luce, 1986].

When participants listen for a target phoneme in nonsense words, they compare the incoming signal with their mental representation of the target. Naturally, languages differ in their exact acoustic manifestation of the phonemes, and, as a consequence, if participants listen to words produced by a speaker of a foreign language, they will probably hear not the best examples of their speech sound categories. Nonetheless, they will extract acoustic cues which are relevant for the identity of the speech sound, and will rely on general acoustic cues to this segment [according to Stevens', 2002, acoustic landmarks], in addition to selecting cues in a language-specific way [e.g., Iverson et al., 2003; Wagner et al., 2006]. Importantly, in contrast to discrimination experiments, in phoneme monitoring experiments listeners are asked to assign the auditory stimulus to a mental representation as fast as possible. In such speeded categorisation

Table 1. Phonemic categories for vowels, stop consonants, and fricatives in the five languages tested

	Vowels	Stop consonants	Fricatives
Catalan	i e ε a ə ɔ o u (8)	p b t d k g (6)	f s z ʃ ʒ (5)
Dutch	i y ɪ e ʏ ø ε ə u o ɔ a ə i æ y ɔ u (16)	p b t d k (5)	f v s z x h (6)
English	ɪ ɪ e æ ʌ u ʊ ɔ ʌ ɒ eɪ ɜ aɪ aʊ ə ə ɔɪ ɪə eə əʊ ʊə (20)	p b t d k g (6)	f v θ ð s z ʃ ʒ h (9)
Polish	ɪ ε a ɪ ɔ u ɛ̃ ɔ̃ (8)	p b t d k g (6)	f v s z ʃʲ ʒʲ ɕ ʑ ʂ ʐ x (11)
Spanish (Castilian)	ɪ e a o u (5)	p b t d k g (6)	f θ s x (4)

The numbers of categories are given in parentheses.

tasks, listeners’ reaction times have been shown to be hardly affected by goodness of stimulus category [Flege et al., 1994; Miller 2001].

Experiment

Languages Compared

We compared listeners of five different languages: one Slavic language (Polish), two Germanic languages (Dutch and British English), and two Romance languages (Catalan and Castilian Spanish). Among the many differences between these languages, the focus in this study is on the numbers of categories for the three speech sound classes – vowels, fricatives, and stop consonants. Table 1 displays the phonemes in these classes in the five languages.

Dialectal variations within a language add or eliminate some phonemes for certain listener groups. Also, due to language-specific phonotactic rules, phonemes may differ in their frequency of occurrence, and their occurrence may be restricted to certain contexts. For instance, Spanish listeners acquire four different fricatives in their native language, but one of them, the /x/, seldom occurs in word-final position [e.g., see LEXESP, Sebastián-Galles et al., 2000]. Furthermore, phonological descriptions of even the same language variety may list different numbers of phonemes. The numbers in table 1 can be considered as averages of the proposed numbers and as the numbers that most authors agree on. We followed Carbonell and Llisterri [1992] for Catalan, Martinez-Celdran et al. [2003] for Castillian Spanish, Booij [1995] for Dutch, Ladefoged [2001] for British English, and Rothstein [1993], and Zygis and Hamann [2003] for Polish.

For some languages, the vowels include diphthongs. The definition of diphthong has been subject to a debate among phoneticians for decades [cf. Gottfried et al., 1993]. In the present study, only diphthongs which are consistently described as consisting of two vowel qualities were taken into account. Hence, we counted diphthongs as different vowel categories only for Dutch and British English [Booij, 1995; Fry, 1979;

Ladefoged, 2001; Rietveld and Van Heuven, 2001, p. 71]. Some descriptions of the phoneme inventories of Spanish and Catalan also contain the notion of diphthongs, but these diphthongs are formed by one of the glides /j/ or /w/ and a vowel [e.g., Green, 1990; Martinez-Celdran et al., 2003]. For the same reasons the British English diphthong /ju/ as in *hue*, was not counted as a vowel category for English.

The variation in the numbers of categories among the languages is evident. For instance, if we consider the fricatives, we see that Polish listeners discriminate nearly twice as many categories as Catalan, Dutch or Spanish listeners. With respect to the vowel categories, British English listeners distinguish approximately four times as many vowels as Spanish listeners. The smallest variation among the languages appears in the distribution of stop consonants. The number of categories is treated as an independent variable in the analyses, and is given in parentheses in table 1.

Naturally, these languages also differ in the exact realisation of the phonemes. For instance, Spanish and Catalan speakers produce stop consonants without aspiration, Dutch and Polish speakers with little aspiration, and English speakers with long aspiration following the burst. Similarly, the vowels in these languages differ in their average formant values [see, e.g., Bradlow, 1995, and the chapters on the relevant languages in International Phonetic Association, 1999]. The fricative targets in the present study (/s/ and /f/) show the least variation among the standard variants of the languages tested. For a more detailed description of the acoustic properties of fricatives in these languages see Jongman et al. [1998] for English, Rietveld and Van Heuven [2001] for Dutch, Jassem [1965] for Polish, and Borzone and Massone [1981] for Spanish. Note that, however, as described above, phoneme monitoring will hardly be affected by variation at this low phonetic level.

Materials

We created 60 words consisting of three, and 60 words consisting of four consonant-vowel (CV) syllables. The consonants were of the set /p t k f s/, and the vowels of the set /a i u o e/. Each phoneme occurred only once per word. These CV strings were nonsense words in all the languages tested.

In these 120 critical items, the target phonemes, /p t k f s a i u/, were always in the final syllable (e.g., /p/ or /u/ could be the target in *fasipu*). Each consonant appeared as target in 15 nonsense words, forming a syllable with each of the three vowels /a i u/ in five nonsense words. Similarly, each vowel appeared as a target in combination with one of the consonants /p t k f s/ in three nonsense words. Appendix A lists all the critical items and the corresponding target phonemes.

In addition to these critical items, 15 nonsense words were created for each target phoneme in which the target appeared in the penultimate syllable, and 15 nonsense words in which the target was missing. Ten practice items were created as well, which familiarised listeners with the experimental situation before the actual test period started.

A male Spanish speaker read the list of stimuli with primary stress on the first syllable. He was instructed to produce the words as if they were existing Spanish words. Thus, the plosives in the materials were unaspirated, the vowels were produced according to Spanish qualities and quantities, and the fricatives were labiodental /f/, and apical alveolar /s/. Recordings were made in a sound-attenuated room directly to a computer, and then down-sampled to 22.05 kHz (16 bit resolution).

Procedure

Participants sat in a sound-attenuated room in front of a computer screen. They were presented with the stimuli over headphones. The trials were blocked by target phoneme, with the order of blocks

counterbalanced among participants. Every block of stimuli was followed by a break, the duration of which was controlled by the participants themselves.

The experimenter informed all participants orally about all targets before the experiment started. In addition, during the experiment a letter appeared on the computer screen designating the current target sound. The English listeners also heard this target over their headphones at the beginning of the block, because of the large grapheme-phoneme discrepancy in English. These auditorily presented phoneme realisations were recorded by a phonetically trained speaker, who produced the targets following the phrase 'Press the button as soon as possible when you hear an...'. The target phonemes were realised as a labiodental [f], an alveolar [s], unaspirated stop consonants, and the vowels [a], [i], [u]. The speaker produced the vowels as close as possible to their cardinal positions. A small group of native listeners of the languages tested judged that these vowels sounded like good examples of vowels in their language.

Participants were instructed to press a key as soon as they recognised the target phoneme in the aurally presented materials. From the onset of each item, listeners had 2,000 ms to respond. Failures to respond, and response latencies over 2,000 ms, were defined as timeout errors. The experiment was self-paced: the next stimulus was presented 1,000 ms after the participant's response or, in case of a timeout, 3,000 ms after the onset of the previous trial, and it was preceded by a beep tone.

For the analyses, we measured the reaction times from the onsets of the target sounds. These onsets were determined visually on the basis of the waveform and spectrogram of the signal. For the vowels, the onset was defined as the onset of voicing. For fricatives, the onset was the offset of voicing in the preceding vowel. The onset of stop consonants is more difficult to define. In previous studies the onset was defined as the onset of the burst [but cf. Cutler and Otake, 1994]. There are, however, reasons to measure reaction times from closure onset, as the closure itself is a cue to manner and as the preceding vowel provides information about place of articulation. By measuring the reaction times from closure onset, a fairer comparison is possible between stop consonants and fricatives, which are also measured from a point directly following the formant transitions in the preceding vowel. In the present study reaction times were therefore measured first from the onset of the closure. In supplementary analyses, we included reaction times measured from the release burst in order to compare our data with previous results.

Participants

Twelve native Dutch speakers were recruited from the subject pool of the Max Planck Institute in Nijmegen. In addition, 12 Spanish native speakers who were spending an exchange period in Nijmegen participated in this experiment. Furthermore, 9 Catalan listeners were tested at the Universidad de Barcelona, 12 native speakers of Polish at the Uniwersitet Śląski in Katowice, and 12 native speakers of British English at the University of Sussex in Brighton, UK. Care was taken that the listener groups were as homogenous as possible with respect to dialectal background. In particular, only those Spanish exchange students were recruited whose native dialect did not belong to the group of dialects spoken in Catalonia. None of the participants reported any speech or hearing disorders. Their participation was rewarded with a small amount of money or with credits needed for their studies.

Results

Reaction Times

Reaction times shorter than 100 ms and longer than 1,500 ms were excluded from the analysis (0.8% of the data). Table 2 shows the mean reaction times for the three phoneme classes, and the five listener groups.

One way to analyse the reaction times would be to just compute the averages for the different languages and phoneme classes and analyse these averages for effects of phoneme class and number of categories. Such an analysis, however, would not be very reliable. The averages would not only reflect the effects of phoneme class, number of

Table 2. The average response times (in milliseconds) for the three phoneme classes and the five languages

	Vowels	Stop consonants	Fricatives
Catalan	435	480 (396)	500
Dutch	475	522 (436)	442
English	524	554 (467)	538
Polish	589	626 (540)	648
Spanish	557	655 (570)	635

For stop consonants reaction times were measured from both the onset of the closure (first number) and from the onset of the release burst (in parentheses).

categories, structural differences between the languages (e.g., syllable structure and stress patterns), but also reflect differences between the average speeds of the different groups of participants resulting from their familiarity with the experimental task.

Instead of comparing the average reaction times of the different language groups for the three phoneme types, we analysed the data by means of multilevel regression models [e.g., Venables and Ripley, 2002; Baayen, in press]. We inserted Language, but also Participant and Item, as crossed random effects. This implies that the model computes different intercepts for each combination of language, participant and item. In other words, it partials out the effects of these factors while computing the effects of the fixed predictors of interest. This enormously reduces the variance in the data. As a consequence, this model is able to detect patterns in the data that are not easily visible in simple scatter plots. Moreover, the inclusion of Language, Participant, and Item as random effects allows us to generalize the observed effects of the fixed predictors over languages, listeners, and words.

The two main variables of interest are the Phoneme Class of the target and the Number of Categories in its class in the participant's language. We considered the log of the Number of Categories, instead of the bare number of categories, since preliminary analyses showed a non-linear relation between the reaction times and the number of categories. However, Phoneme Class of a target predicts its number of Categories to some extent, in particular for the plosives (table 1). The two predictors are thus col-linear and just entering both of them into the model may lead to misleading results [cf. Chatterjee et al., 2000]. We therefore orthogonized the two variables as follows. We ran a simple linear model predicting the log number of Categories as a function of Phoneme Class. The residuals of this model are highly correlated with the log Number of Categories ($r = 0.829$, $p < 0.0001$), but display no relationship with Phoneme Class. We entered these residuals (henceforth: Residuals of the Number of Categories, RNC) together with Phoneme Class as fixed effects in the multilevel regression model for the reaction times. A potential interaction between Phoneme Class and RNC was excluded from the initial model, as it could not provide meaningful results: the languages hardly differ in their numbers of categories for stop consonants, while they differ strongly in their number of vowels.

Table 3 lists the statistics for this initial model. Both Phoneme Class, [$F(2, 5,636) = 23.75$, $p < 0.001$] and RNC, representing the Number of Categories [$F(1, 5,636) = 27.80$, $p < 0.001$], were significant. Additional analyses showed that participants' reactions were significantly faster to vowels (mean reaction time: 526 ms) than to fricatives [565 ms, $F(1, 3,529) = 19.19$, $p < 0.001$] and stop consonants [577 ms, $F(1, 4,075) =$

Table 3. Estimated values for the fixed effects and the random effect of Language in the model for the reaction times

Fixed effects	
Intercept (Fricative, number of categories = 0)	564.81
Stop consonant	14.57 (−71.01)
Vowel	−38.86
RNC	45.86 × RNC
Random effect of Language	
Catalan	−49.67
Dutch	−56.67
English	−7.84
Polish	48.89
Spanish	65.30
Degrees of freedom	5,639

For stop consonants, the first number refers to the reaction times measured from the closure onset, while the number in parentheses refers to the measurements from onset of the release burst.

40.46, $p < 0.001$], which did not differ from each other ($p > 0.05$). The effect of RNC showed that a higher number of categories slowed down listeners' responses.

Figure 1a shows the modelled relation between the plain number of categories and the reaction times for a listener monitoring a fricative (only the intercept changes for vowels or stop consonants). The relation is non-linear and shows attenuation of the effect of the number of categories at higher numbers: an additional phonemic category has a bigger impact on the response latencies if only few categories are present in the phoneme class than if there are already many categories. Note that this significant positive relationship between response latencies and number of categories is not obvious from the average response times listed in table 2. The reason for this is that the participants of the five languages differed in their average reaction times. The model partials out this variance by means of the random effects of Language and Participant.

To examine a possible difference in the effect of the number of categories on the identification of fricatives and vowels (the languages tested do not differ in the number of categories for stop consonants), we conducted a second analysis. Again, we modelled the reaction times as a function of Phoneme Class and RNC, but now also included a potential interaction between these two factors. As in the first analysis Language, Participant, and Item were included as crossed random factors. Stop consonants were excluded from this analysis. This second analysis showed significant main effects of Phoneme Class [$F(1, 3,528) = 19.03$, $p < 0.001$] and RNC [$F(1, 3,528) = 27.14$, $p < 0.001$]. The interaction between Phoneme Class and RNC also emerged as significant [$F(1, 3,528) = 5.57$, $p < 0.05$]. Further analyses showed that RNC affects both classes, but the effect is bigger for the vowels than for the fricatives.

In the analyses reported above, the reaction times to stop consonants were measured from closure onset. This implies that the reaction times include a period of silence that precedes the crucial cues carried by the release burst. Moreover, our data now cannot be directly compared with the results of previous studies [e.g., Foss and Swinney, 1973; Savin and Bever, 1970]. We therefore also ran the analyses with the reaction times for the stop consonants measured from burst onset. Again we excluded reaction times shorter than 100 ms and longer than 1,500 ms. The analyses show again an effect of RNC [$F(1, 5,635) = 28.53$, $p < 0.001$], with a higher number of categories

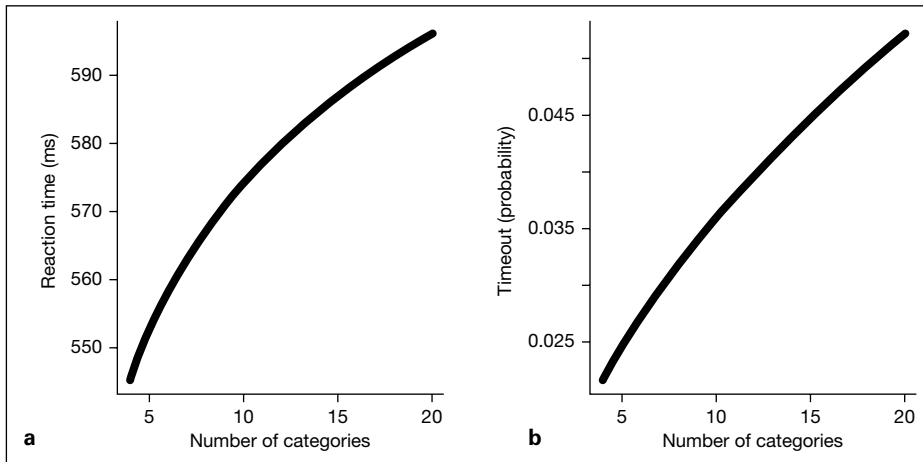


Fig. 1. The modelled relation between the plain Number of Categories and Reaction Times (**a**) and the probability of a timeout error for a listener detecting a fricative (**b**).

leading to slower responses. Phoneme Class [$F(2, 5,635) = 32.61, p < 0.001$] was also significant. Reactions to stop consonants were now the fastest [mean reaction time for stops: 492 ms, difference with vowels: $F(1, 4,075) = 14.88, p < 0.001$; difference with fricatives: $F(1, 3,667) = 77.78, p < 0.001$]. This is as expected, since the reaction times to stop consonants were on average 80 ms shorter than in the previous analysis.

Errors

Table 4 displays the absolute numbers of timeouts and non-timeouts for the phoneme classes and for the five language groups. The percentages of timeouts are given in parentheses. We modelled the probability of a timeout with a generalised multilevel model, with Language, Item, and Participant as random factors. The predictors considered as fixed effects were Phoneme Class, and RNC, representing the number of categories.

Both RNC [$F(1, 6,164) = 33.38, p < 0.001$] and Phoneme Class [$F(2, 6,164) = 10.14, p < 0.001$] appeared significant (see table 5 for the effect sizes). Further analysis showed that listeners missed more vowels than fricatives [$F(1, 3,895) = 16.17, p < 0.001$] or stop consonants [$F(1, 4,498) = 12.09, p < 0.001$], but showed no difference between these two latter classes ($p > 0.1$). To illustrate the effect of the number of categories, figure 1b shows the predicted probability of a timeout error for participants monitoring fricatives as a function of the plain number of categories. For the other phoneme classes only the intercept changes.

Table 5 also shows the random effects of the languages. The languages clearly differed in their mean percentages of timeouts. For instance, the English listeners missed more targets than the Catalan listeners. One reason for this may be that the listener groups differed in their familiarity with the experimental task. The English participants, for instance, had less experience with psycholinguistic experiments than the Catalan participants. Another reason may be that the pronunciation of the Spanish speaker is more native-like to the Spanish and Catalan listeners than to the other listener groups. We investigated potential effects of speaker in a control experiment.

Table 4. The absolute numbers of timeouts and non-timeouts for the three phoneme classes and the five languages

	Vowels	Stop consonants	Fricatives
Catalan	6/238 (2.46%)	1/238 (0.42%)	3/177 (1.67%)
Dutch	46/422 (9.83%)	24/463 (4.93%)	13/347 (3.62%)
English	104/428 (19.55%)	60/473 (11.26%)	45/332 (11.94%)
Polish	26/454 (5.42%)	7/473 (1.46%)	19/341 (5.28%)
Spanish	48/459 (9.47%)	66/565 (12.43%)	20/370 (5.13%)

The percentages of timeouts are given in parentheses.

Table 5. Estimated values for the fixed effects and the random effect of language in the model for the timeout errors

Fixed effects	
Intercept (Fricative, number of categories = 0):	−3.46
Stop consonant	0.14
Vowel	0.60
RNC	0.82 × RNC
Random effect of Language	
Catalan	−1.59
Dutch	0.06
English	0.54
Polish	−0.34
Spanish	0.82

Control Experiment

In the main experiment all listeners were presented with the same materials realised by a Spanish speaker. The Spanish participants were thus listening to native realisations of the nonsense words, whereas the other participants heard foreign pronunciations. It has been shown that when listening to speech on a low phonetic level, as in phoneme monitoring, listeners apply their native listening strategies, which are defined by their phonology and their phoneme inventories [Costa et al., 1998] and are hardly affected by the exact acoustic realisation of the materials [e.g., Cutler and Otake, 1994; Wagner et al., 2006]. Nevertheless, it is possible that the materials were processed in different ways by native and non-native listeners.

In order to test whether the effects of the class of the phoneme and the number of categories are present independently of the precise acoustic realisations of the stimuli, we ran a control experiment. In this experiment, Spanish and Dutch listeners were presented with materials realised by a native speaker of Dutch.

Materials, Procedure and Participants

A native speaker of Dutch recorded the experimental stimuli, in addition to some new fillers. The materials were very similar to those in the main experiment, but lacked stimuli with /k/ as a target. The Dutch speaker was asked to produce good examples of the Dutch phonemes. Thus, the Dutch stop consonants were realised with a short period of aspiration, the target vowels sounded like Dutch phonetically short /i u/ or long /a/, the fricatives were again the labiodental /f/, and alveolar /s/. Recordings

Table 6. The mean reaction times (RT) and the absolute numbers of timeouts and non-timeouts for the Spanish and Dutch listeners in the control experiment with a Dutch speaker

		Vowels	Stop consonants	Fricatives
Dutch	RT	505	560	492
	Errors	39/311 (11.14%)	14/236 (5.6%)	7/223 (3.04%)
Spanish	RT	552	613	625
	Errors	23/327 (6.57%)	17/233 (6.8%)	12/228 (5%)

Reaction times to stop consonants were measured from the onset of the closure. The percentages of timeouts are given in parentheses.

were made in a sound-attenuated room directly to a computer, and then down-sampled to 22.05 kHz (16 bit resolution). The procedure of the experiment was as in the main experiment.

Ten new Dutch speakers from the subject pool of the Max Planck Institute, and 10 new Spanish exchange students in Nijmegen were recruited to take part in the control experiment. None had participated in the main experiment, and none had any known speech or hearing disorders.

Results

Table 6 presents the average reaction times and the numbers and percentages of timeouts for these two new groups of participants, broken by Phoneme Class.

Reaction Times

Reaction times to stop consonants were analysed as in the main analysis, thus measured from closure onset. The data from the main experiment were pooled with the data from the control experiment. We then analysed the data for all Spanish listeners for an effect of Speaker. We entered Speaker together with Phoneme Class as fixed effects in a multilevel regression model for the reaction times, with Item and Participant as crossed random factors. Note that we could not investigate the effect of the number of categories in this analysis, since this number is completely predictable given the class of the phoneme (as there is only one language). The effect of Phoneme Class emerged as significant [$F(2, 2,069) = 33.78, p<0.001$]: Spanish listeners identified vowels significantly faster than stop consonants [$F(1, 1,476) = 76.26, p<0.001$] and fricatives [$F(1, 1,374) = 34.41, p<0.001$], while there was no difference between stop consonants and fricatives ($p>0.1$). More importantly, the effect of Speaker was not statistically significant, neither was its interaction with Phoneme Class ($p>0.1$). We then performed the same analysis for the two groups of Dutch participants and attested an effect of Phoneme Class [$F(2, 1,987) = 13.28, p<0.001$]. Both Dutch groups identified vowels faster than stop consonants [$F(1, 1,420) = 14.25, p<0.001$]. Fricatives were also identified significantly faster than stop consonants [$F(1, 1,265) = 30.46, p<0.001$], but there was no difference between vowels and fricatives ($p>0.1$). Importantly, also this analysis showed no significant effect of Speaker ($p>0.1$) and no significant interaction ($p>0.1$). These data suggest that the Spanish and Dutch listeners were not affected by whether they were familiar with the exact acoustic realisations.

We also ran another analysis, which addresses more directly whether the effect of the number of categories is robust against different acoustic realisations. In this analysis the data for the Spanish and Dutch participants from the main experiment were removed

from the data set and we only kept the data from the Spanish and Dutch participants in the control experiment. Reaction times were modelled as depending on RNC and Phoneme Class, with Language, Participant and Item as cross-random factors, as we did for the data of the main experiment. Both RNC [$F(1, 4,669) = 4.24$ $p = 0.03$] and Phoneme Class [$F(2, 4,669) = 17.20$, $p < 0.001$] were again significant, showing that both effects are robust and that the exact realisations of the materials are not decisive.

Errors

We analysed the errors of the control experiment in the same steps as we analysed the reaction times. The analysis of all Spanish listeners as well as the analysis of all Dutch listeners showed neither a main effect of Speaker nor any interaction with Speaker. The analysis of the data set of the main experiment with the Spanish and Dutch listeners replaced by the Spanish and Dutch listeners from the control experiment revealed main effects of both Phoneme class [$F(1, 5,089) = 11.18$, $p < 0.001$] and RNC [$F(1, 5,089) = 19.84$, $p < 0.001$]. Participants made more errors for vowels and more errors if the number of categories in the phoneme's class was higher. In conclusion, the control experiment shows that the effects of Phoneme Class and Number of Categories are inherent to phoneme monitoring and independent of whether the listeners hear a native or a non-native pronunciation of the nonsense words.

Phoneme Frequencies

The correlation of the reaction times and the timeout errors with RNC suggest that the speed and ease of phoneme identification depend on the sizes of listeners' phoneme repertoires. However, there may be an alternative explanation for our results. The targets in our experiments are the most common phonemes in the world's languages, but the relative frequencies of occurrence of these phonemes vary across languages. Importantly, a language with more categories in its phoneme inventory may make less use of the phonemes tested in our experiments. In other words, there may be a confound between the number of categories and the frequencies of occurrence of the phonemes in the languages. Hence, the attested effect of the number of categories might actually be an effect of frequency of occurrence, and listeners with a smaller number of categories may be faster and more efficient in identifying phonemes just because of the more frequent occurrences of these phonemes in their language.

One might assume that listeners are so proficient in recognising their native phonemes that their performance is at ceiling, and that frequency cannot influence their performance in phoneme monitoring. Nevertheless, there are results pointing in the direction that phoneme frequency does play a role in phoneme identification. Warner et al. [2005] examined the effects of phoneme frequency on listeners' guesses about the identity of a segment in a gating study where listeners heard increasing portions of phonemes in a random order. A correlation was observed between phoneme frequency and listeners' decisions, when little acoustic information about the segment was available (that is, at short portions of the signal). For longer portions this correlation decreased gradually. Hence, faster and more accurate identifications may be due to higher phoneme frequencies, instead of lower numbers of categories, in the present task as well.

To investigate this issue, we carried out two types of analyses. First, we examined whether the number of categories within a phoneme class is correlated with the

frequencies of occurrence of its phonemes. Second, we reanalysed the response latencies and timeout errors including frequency as an additional predictor.

From the set of the languages tested, phoneme frequencies could be determined for Dutch, English and Spanish, as phonemically transcribed databases of words are available for these languages, which also include information about the token frequencies of the words [CELEX for Dutch and English, see Baayen et al., 1993; LEXESP for Spanish, see Sebastián-Galles et al., 2000]. We calculated the frequencies of occurrence of the phonemes per million phonemes, taking into account the token frequencies of the words (token frequency) or just counting every word once (type frequency).

We computed the correlation of the log number of categories for the phonemes with their log frequencies in the three languages. We found no correlation with the log token frequency of the phonemes. However, the log number of categories appeared highly correlated with the log type frequency ($r = -0.54$, $p < 0.01$). Unsurprisingly, listeners with fewer categories in their native phoneme repertoire make more frequent use of these phonemes in the words in their vocabulary.

In the second analysis, we included the log token and type frequencies of the phonemes as additional predictors in our model for the response latencies, for the three languages for which frequency values were available. In this analysis, the log phoneme frequencies were no significant predictors. The variance in the reaction latencies is explained by Phoneme Class and RNC but not by phoneme frequencies.

For the timeout errors, however, we observed a main effect for the log token frequency of the phonemes. A higher phoneme frequency implied fewer errors [$F(1, 4,180) = 4.33$, $p < 0.05$]. Importantly, the main effect of RNC was still significant [$F(1, 4,180) = 22.26$, $p < 0.001$]. This is as expected, as RNC, reflecting the log number of categories, was not correlated with the log token frequency of the phonemes.

In conclusion, the attested effect of the number of categories is not a frequency effect in disguise. In addition, our results support the view that phoneme monitoring may be affected by the frequencies of occurrence of the phonemes. However, the effect appears to be limited to participants' accuracy and not to extend to response latencies.

General Discussion

This study investigated how listeners' speed and accuracy in phoneme identification is affected by the class of the speech sound (vowel, stop consonant, fricative) and by the number of categories within this class in the listeners' native phoneme repertoire. In a phoneme monitoring experiment with nonsense words, native listeners of five different languages (Castilian Spanish, Catalan, Dutch, British English, and Polish) identified vowels, fricatives, and stop consonants that represent phonemes in all the five languages. The results show that listeners identified vowels more quickly than fricatives. There was no difference between fricatives and stop consonants if reaction times to stop consonants included the interval of the closure. If the reaction times were measured from burst onset, however, as in previous studies, stop consonants were identified more quickly than fricatives and vowels. Phoneme class also affected participants' accuracy: consonants were identified more accurately than vowels. Furthermore, we found an effect of the number of categories: a phoneme is recognised faster and more accurately if it has fewer competitors belonging to the same class in the listener's phoneme inventory.

The present study extends previous research on differences between phoneme classes to more languages. Whereas nearly all previous findings are based on English [e.g., Foss and Swinney, 1973; Morton and Long, 1976; Healy and Repp, 1982; van Ooijen, 1994], we studied listeners of Romance languages, Germanic languages, and a Slavic language. Our study replicates the finding that stop consonants, with reaction times measured from burst onset, are recognised faster than fricatives [Foss and Swinney, 1973; Morton and Long, 1976; Rubin et al., 1976].

Several studies, summarised in the 'Introduction', have attributed this difference between fricatives and stop consonants to mechanisms of auditory processing. Stop consonants would be processed more categorically: due to their acoustic properties, their perceptual traces would decay faster, such that their recognition would be mainly based on traces in phonetic memory. Fricatives, on the other hand, would be perceived more continuously and processed on the basis of the more detailed traces in the auditory memory. As a consequence, stop consonants may be labelled faster than fricatives.

Our results show that there is an alternative explanation, which lies in the decision about the onset of the measurements of the reaction times. Stop consonants consist of the silent interval of the closure and of the abrupt release burst. Most previous studies have measured the response latencies for stop consonants from the release burst. However, the silent interval of the closure provides cues to the manner of articulation of the consonant and its duration may provide information about place of articulation. Moreover, the onset for reaction times for fricatives is set immediately after the formant transitions in the preceding vowel. By measuring the reaction times for stop consonants from the release burst, that is, much later than the end of the formant transitions, there is no fair comparison possible between fricatives and stop consonants. Obviously, a conclusion about which phonemes are identified more slowly depends very much on the onset of measurement for the reaction times. If we measure from closure onset, we see that labelling phonemes based on phonetic memory (stop consonants) or auditory memory (fricatives) does not necessarily lead to differences in identification times.

We also found that fricatives were in general identified more slowly than vowels. This seems to be in contrast to the findings of Savin and Bever [1970] and van Ooijen [1994], who reported that vowels are detected more slowly than fricatives for Dutch and English. This contrast is only apparent. Table 2, listing the average reaction times for the different languages and phoneme classes, shows that also in our experiment, Dutch listeners detected vowels more slowly (mean reaction time: 475 ms) than fricatives (442 ms) and that there is hardly any difference between the two phoneme classes for the English listeners tested. Catalan, Polish and Spanish listeners, on the contrary, recognised vowels more quickly than fricatives. These differences between listener groups demonstrate the effect of the numbers of categories within the three phoneme classes that substantially vary among the languages tested (see below). After the effect of the number of categories is partialled out, vowels were in general recognised faster than fricatives.

In the 'Introduction', we formulated a hypothesis about ease of identification of vowels versus consonants on the basis of their function in lexical processing. Since vowels have been shown to constrain lexical selection to a lesser extent [e.g., Cutler et al., 2000], they might also be identified more slowly and less accurately. Regarding the accuracy of identification we found that listeners indeed made more errors on vowels than on consonants. Regarding the response latencies, however, we found exactly the opposite of what we predicted: vowels were identified more quickly than consonants. One possible explanation may be that participants were less cautious in their reactions

to vowels, exactly because vowels restrict lexical selection to a lesser extent than consonants. This would lead to faster responses but also to more errors.

Another explanation for the fast responses to the vowels may lie in their acoustic manifestation. More acoustic cues are present in the preceding context for vowels than for fricatives. For instance, whereas the formant transitions following consonants are generally assumed to be perceptually more relevant than the preceding formant transitions [e.g., Stevens and Blumstein, 1978], important cues for the identity of the vowel are present in the preceding consonant [e.g., Whalen, 1981] and even the preceding vowel [e.g., Manuel, 1990]. This acoustic difference between vowels and fricatives may be determinative and interfere with any other effects.

We now turn to the role of the number of categories that we have documented for phoneme monitoring. A higher number of categories slowed participants down and made them less accurate. Since the number of categories within a phoneme class is language-specific, its effect yields language-specific patterns in phoneme recognition.

Table 2 shows that there were roughly three rankings of the phoneme classes among the five languages. First, Catalan, Polish, and Spanish showed the basic pattern which emerged from our statistical analyses with number of categories as a predictor: Vowels were recognised faster than fricatives and stop consonants (as measured from closure onset). Second, the English participants recognised vowels as slowly as fricatives and stop consonants. This pattern is in line with the high number of vowels in this language, which makes listeners recognise them more slowly. Finally, in Dutch, vowels were recognised slightly more slowly than fricatives. This is in line with the high number of vowels also in this language, in combination with a relatively low number of fricatives.

These results illustrate that cross-language research is necessary to gain insight into speech processing. Studies investigating only one language tend to attribute differences between phoneme classes to the acoustic properties of the speech segments. This however is not the only factor contributing to listeners' phoneme identification, since, as we have shown, a major factor is the number of categories in the listener's native language. This factor can only be documented by comparing several languages.

Interestingly, the effect of the number of categories appeared to be greater for vowels than for fricatives (and possibly stop consonants). One explanation for this difference is in line with the fast responses and lower accuracy that we attested for vowels (see above). Vowels restrict lexical selection to a lesser extent, therefore listeners may generally pay less attention to vowels, and as a consequence be more sensitive to factors inhibiting identification.

The effect of the number of categories may stem from general properties of the perception system. A higher number of categories within a class implies a higher number of choices, which generally impedes the process of decision making [e.g., Medin et al., 1995; Nosofsky, 1997]. This holds especially if the choice options are highly similar [Foss and Dowell, 1971]. For instance, in visual perception, the search for an object on a display is slowed down both by a higher number of alternatives (set size effect) and a greater similarity among these alternatives [Theeuwes, 1992, Palmer et al., 2000].

Note that in the experiments in the visual domain, the alternatives are in general all present on the display, and therefore their number and similarity can be manipulated within participants. In phoneme monitoring, the alternatives are the categories in the participants' native phoneme repertoires. All these categories affect participants' phoneme monitoring, even though they are not all incorporated in the materials of the

experiment. In consequence, number and similarity of categories cannot be manipulated within participants, and need cross-linguistic investigation.

The effect of the number of categories can be explained by several aspects of identification processing. First, a greater set of phonemes involves the exclusion of more potential candidates in the search for a mental representation to match the presented signal. This implies a greater combined probability of incorrect candidates. Second, a higher number of categories implies that the perceptual space will contain more boundaries, and more sounds will be positioned at boundaries. In consequence, more sounds may be ambiguous and might therefore be harder to classify. Third, a greater set of phonemes implies that more phonemes share acoustic features, and fewer features distinguish a phoneme from its competitors. According to several categorisation models [e.g., Ashby, 2000; Nosofsky, 1997], the degree of similarity between the alternatives affects reaction times and categorisation accuracy. Hence, both the number of categories in a class and the similarity between speech sound categories may have contributed to our results.

The perceptually similar categories are not necessarily phonemes sharing the manner of articulation (phoneme class), but may also share other acoustic features. For instance, the voiced bilabial stop consonant /b/ may be perceptually similar to the voiced bilabial fricative /v/. Further research is necessary to determine the precise effects of the number of categories in a phoneme class and the numbers and types of similar phonemes in the language belonging to different classes. Note that for such research, the degree of similarity between every pair of phonemes needs to be established separately for each language, as this degree might not only be determined by the phonemes' acoustics, but also by the phonotactic constraints in the language.

Furthermore, in the present study, we have made the simplified assumption that the categories that affect speech identification represent different phonemes. In line with this assumption, the numbers of categories that we used in our statistical analyses are the numbers of native phonemes in the different classes. However, listeners are also capable of discriminating allophonic variants of the same phoneme [e.g., Lipski, 2006]. Future research has to show whether the number of distinctive speech sounds in a class is a better predictor than the number of phonemes. Note that this research will only be possible once we know which speech sounds listeners of the different languages can distinguish.

The study by Costa et al. [1998] shows that in phoneme monitoring listeners are aware of the acoustic variation in the realisation of a phoneme that can be induced by co-occurring native phonemes. Our study shows that in addition listeners are aware of the acoustically similar phonemes in their native language. Apparently, listeners' identification of phonemes is affected by the acoustic variability within the category of a phoneme, and also by the number of categories within the phoneme's class.

Interestingly, no difference was found between participants listening to a native or to a non-native speaker. This result shows that when listening to phonemes in nonsense words, a situation which resembles listeners' first contact with a foreign language, listeners assimilate speech sounds to their own native categories. The exact acoustic realisation of a speech sound which is phonemic in the native language hardly affects listeners' identification.

The effect of the number of categories is also present in subsets of the data. It is also significant if we do not take into account the stop consonants or exclude one of the five languages (e.g., Spanish or English). Probably, the effect would have been even stronger if the languages had been more similar in their syllable structure, stress patterns, phonotactic constraints, etc. Of course it is impossible to control for such differences

between languages. The robustness of the effect of the Number of Categories suggests that this effect is inherent to phoneme monitoring.

Importantly, the effect of number of categories cannot be explained by the frequencies of occurrence of the phonemes in the respective languages. As may be expected, in the languages for which frequency counts are available (Dutch, English, and Spanish), the frequencies of the phonemes in the languages' vocabularies are negatively correlated with the numbers of categories in their classes. Thus, languages with fewer phonemes in a given class use these phonemes more frequently in their words. We investigated whether the frequencies of the phonemes were predictors for the response latencies for the Dutch, English, and Spanish participants, in addition to the number of categories, but this was not the case. For the accuracy, we also found that incorporating the frequencies of the phonemes did not reduce the effect of the number of categories. Hence, the effect of the number of categories is not a frequency effect in disguise.

Nevertheless, we observed that phoneme frequency played a role in participants' accuracy: participants made fewer errors for phonemes that occur more often in their speech (word tokens). However, the role of frequency appears minor as it only surfaces in the number of errors.

Given that higher numbers of categories in phoneme classes slow listeners down, one might expect that small numbers of categories form a preferable pattern among the languages of the world. Indeed, despite the great variation in the number of phonemes in the world's phoneme inventories, ranging from 11 (Rotokas) to 141 (!Xu), more than 70% of languages have between 20 and 37 segments [Maddieson, 1984]. In natural speech interactions listeners' purpose is not to identify phonemes, but to recognise words to apprehend their meanings. Listeners would be hindered by lower numbers of phonemic categories, as this would lead to longer words, a higher number of words embedded in other words [Cutler et al., 2004], and higher neighbourhood densities, which inhibit lexical access [e.g., Vitevitch and Luce, 1999].

In conclusion, this study documented two sources of variance in phoneme identification which affect listeners of different native backgrounds in the same way: (1) the acoustic and functional properties of the phoneme, and (2) the number of native categories within the phoneme's class. We found these general patterns across five languages, despite the many differences between the languages, for instance, in syllable structure, phonotactic constraints, and stress patterns, which might hide general patterns. The effect of the number of categories in the listener's native phoneme inventory proves to be another consequence of listeners becoming experts in their native phonology. Native speech sounds establish mental references early in speech development, and permanently divide listeners' perceptual space into distinct sound categories. While listeners do not focus on speech sounds in natural speech interactions, individual sounds enter listeners' focus of attention, for instance, when listening to speech in noise, when listening to a speaker with an unusual pronunciation, or when acquiring a foreign language. It may be especially under these conditions that phoneme class and the number of native categories within a class affect speech processing.

Acknowledgements

We are grateful to Prof. Alan Garnham from the University of Sussex, to Dr. Jolanta Tambor from Uniwersytet Śląski in Katowice, and to Prof. Nuria Sebastián-Gallés for helping us in conducting

the experiments in England, Poland, and in Spain. We would further like to thank Anne Cutler, Klaus Kohler, and two anonymous reviewers for their helpful comments on earlier versions of this text. This research was supported by the NWO SPINOZA project ‘Native and Non-Native Listening’.

Appendix A

List of Materials Used in the Experiment

Consonant target	The following context		
	/a/	/i/	/u/
/f/	tekufa	kotafi	tipefu
	tasifa	tusafi	tokafu
	sokifa	pinasafi	posefu
	tilekofa	temupafi	pilotafu
	sinotufa	tenosafi	simokafu
/k/	posika	petuki	pitaku
	tufika	tusaki	sepiku
	pomiteka	palufoki	tenifaku
	finesoka	femoseki	timafeku
	fiselika	temisuki	petisaku
/p/	kesupa	tefupi	kitepu
	sefupa	fusopi	sikapu
	tekipa	talokepi	tafipu
	felukipa	tenasupi	kenosapu
	selukipa	senokapi	kosefipu
/s/	tekusa	pakesi	tepisu
	tikusa	petasi	fekatisu
	pifunesa	tukesi	kopesu
	telikusa	pomekasi	pokefisu
	pilufesa	fukeposi	tilokasu
/t/	tekuta	pakuti	fisetu
	fekuta	kopati	sakitu
	pilefuta	kosati	pemakitu
	fipokuta	pakofuti	felosatu
	simofeta	sekafuti	senoketu

Vowel target	The preceding context				
	/f/	/k/	/p/	/s/	/t/
/a/	petufa	petoka	tofepa	pitosa	pifota
	telisufa	tosuka	fekosipa	fukopesa	pomisuta
	tesupifa	fenusoka	sotipa	tokesa	siputa
/i/	talemofi	telufaki	senufopi	pakotesi	fokesuti
	pasufi	tolepuki	kefopi	fatusi	sefuti
	pomekufi	setuki	tesopi	tukesi	sokati
/u/	tepifu	tomiseku	fekipu	pafisu	fopitu
	sakomifu	paseku	sikapu	fakipesu	finesatu
	somatefu	sutileku	semalipu	tenifasu	pisatu

References

- Ashby, F.G.: A stochastic version of general recognition theory. *J. math. Psychol.* 44 (2000).
- Baayen, H.: Analyzing linguistic data: a practical introduction to statistics (Cambridge University Press, Cambridge, in press).
- Baayen, R.H.; Piepenbrock, H.; van Rijn, H.: The CELEX lexical database (CD-ROM) (Linguistic Data Consortium, University of Pennsylvania, Philadelphia (1993).
- Bonatti, L.L.; Pena, M.; Nespor, M.; Mehler, J.: Linguistic constraints on statistical computations. *Psychol. Sci.* 16: 451–459 (2005).
- Booij, G.: The Phonology of Dutch (Oxford University Press, Oxford 1995).
- Borzone de Manrique, A.M.; Massone, M.I.: Acoustic analysis and perception of Spanish fricative consonants. *J. acoust. Soc. Am.* 69: 1145–1153 (1981).
- Bradlow, A.R.: A comparative acoustic study of English and Spanish vowels. *J. acoust. Soc. Am.* 97: 1916–1924 (1995).
- Bradlow, A.R.: A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish: universal and language-specific aspects. *Phonetica* 53: 55–85 (1996).
- Caramazza, A.; Chialant, D.; Capasso, R.; Miceli, G.: Separable processing of consonants and vowels. *Nature* 403: 428–430 (2000).
- Carbonell, J.F.; Llisterri, J.: Illustrations of the IPA: Catalan. *J. int. phonet. Ass.* 22: 53–56 (1992).
- Chatterjee, S.; Hadi, A.S.; Price, B.: Regression analysis by example (Wiley & Sons, New York 2000).
- Connine, C.M.; Titone, D.: Phoneme monitoring. *Lang. cognit. Processes* 11: 635–645 (1996).
- Costa, A.; Cutler, A.; Sebastián-Gallés, N.: Effects of phoneme repertoire on phoneme decision. *Perception Psychophysics* 60: 1022–1031 (1998).
- Cutler, A.; Otake, T.: Mora or phoneme? Further evidence for language-specific listening. *J. Memory Lang.* 33: 824–844 (1994).
- Cutler, A.; Sebastián-Gallés, N.; Solar-Vilageliu, O.; van Ooijen, B.: Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory Cognition* 28: 746–755 (2000).
- Cutler, E.A.; Mister, E.; Norris, D.; Sebastián-Gallés, N.: La perception de la parole en espagnol: un cas particulier? in Ferrand, Grainger, Psycholinguistique cognitive: essais en l'honneur de Juan Segui, pp. 57–74 (De Boec, Brussels 2004).
- Dupoux, E.; Mehler, J.: Monitoring the lexicon with normal and compressed speech: frequency effects and the prelexical code. *J. Memory Lang.* 29: 316–335 (1990).
- Eisner, F.; McQueen, J.M.: The specificity of perceptual learning in speech processing. *Perception Psychophysics* 67: 224–238 (2005).
- Flege, E.J.; Munro, M.J.; Fox, R.A.: Auditory and categorical effects on cross-language vowel perception. *J. acoust. Soc. Am.* 95: 3623–3641 (1994).
- Foss, D.J.; Dowell, B.E.: High-speed memory retrieval with auditorily presented stimuli. *Perception Psychophysics* 9: 465–468 (1971).
- Foss, D.J.; Swinney, D.A.: On the psychological reality of the phoneme: perception, identification, and consciousness. *J. verbal Learn. verbal Behav.* 12: 246–257 (1973).
- Fry, D.B.: The physics of speech (Cambridge University Press, Cambridge 1979).
- Gottfried, M.; Miller, J.D.; Meyer, D.J.: Three approaches to the classification of American English diphthongs. *J. Phonet.* 21: 205–229 (1993).
- Green, J.N.: Spanish; in Comrie, The major languages of Western Europe (Routledge, London 1990).
- Healy, A.F.; Repp, B.: Context independence and phonetic mediation. *J. exp. Psychol. hum. Perception Performance* 8: 68–80 (1982).
- Hick, W.E.: On the rate of gain of information. *Q. J. exp. Psychol.* 4: 11–26 (1952).
- International Phonetic Association: Handbook of the International Phonetic Association: a guide to the use of the international phonetic alphabet (Cambridge University Press, Cambridge 1999).
- Iverson, P.; Kuhl, P.K.; Akahane-Yamada, R.; Diesch, E.; Tohkura, Y.; Kettermann, A.; Siebert, C.: A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57 (2003).
- Jassem, W.: Formants of fricative consonants. *Lang. Speech* 8: 1–16 (1965).
- Johnson, K.: Massive reduction in conversational American English; in Yoneyama, Maekawa, Spontaneous speech: data and analysis. Proc. 1st Session of the 10th Int Symp., pp. 29–54 (The National International Institute for Japanese Language, Tokyo 2004).
- Jongman, A.; Sereno, J.; Wayland, R.; Wong, S.: Acoustic properties of English fricatives. *J. acoust. Soc. Am.* 103: 3086 (1998).
- Ladefoged, P.: Vowels and consonants (Blackwell, Los Angeles 2001).
- Lieberman, A.M.; Harris, K.S.; Hoffman, H.; Griffith, B.C.: The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.* 54: 358–368 (1957).
- Lipski, S.C.: Neural correlates of fricatives across language boundaries; doct. diss. University of Stuttgart (2006).
- Luce, P.A.: Neighborhoods of words in the mental lexicon. Research on speech perception. Tech. Rep. No. 6 (1986).
- Maddieson, I.: Patterns of sounds (Cambridge University Press, Cambridge 1984).
- Manuel, S.Y.: The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *J. acoust. Soc. Am.* 88: 1286–1298 (1990).

- Martin, R.C.; Breedin, S.D.; Damian, M.F.: The relation of phoneme discrimination, lexical access, and short-term memory: a case study and interactive activation account. *Brain Lang.* 70: 437–482 (1999).
- Martínez-Celdrán, E.; Fernández-Planas, A.M.; Carrera-Sabate, J.: Illustrations of the IPA: Castilian Spanish. *J. int. phonet. Ass.* 33: 255–259 (2003).
- McClelland, J.L.; Elman, J.L.: The TRACE model of speech perception. *Cognit. Psychol.* 18: 1–86 (1986).
- McQueen, J.M.; Cutler, A.; Norris, D.: Phonological abstraction in the mental lexicon. *Cognit. Sci.* 30: 1113–1126 (2006).
- Medin, D.L.; Goldstone, R.; Markman, A.B.: Comparison and choice: relations between similarity processes and decision processes. *Psychonomic Bull. Rev.* 2: 1–19 (1995).
- Miller, J.L.: Mapping from acoustic signal to phonetic category: internal category structure, context effects and speeded categorization. *Lang. cognit. Processes* 16: 683–690 (2001).
- Mirman, D.; Holt, L.L.; McClelland, J.L.: Categorization and discrimination of nonspeech-sounds: differences between steady-state and rapidly-changing acoustic cues. *J. acoust. Soc. Am.* 116: 1198–1207 (2004).
- Morton, J.; Long, J.: Effect of word transitional probability on phoneme identification. *J. verbal Learn. verbal Behav.* 15: 43–51 (1976).
- Näätänen, R.; Lehtokoski, A.; Lennes, M.; Cheour, M.; Huottilainen, M.; Iivonen, A.; Vainio, M.; Alku, P.; Ilmoniemi, R.J.; Luuk, A.; Allik, J.; Sinkkonen, J.; Alho, K.: Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385: 432–434 (1997).
- Norris, D.; McQueen, J.M.; Cutler, A.: Merging information in speech recognition: feedback is never necessary. *Behavioral Brain Sci.* 23: 299–370 (2000).
- Norris, D.; McQueen, J.M.; Cutler, A.: Perceptual learning in speech. *Cognit. Psychol.* 47: 204–238 (2003).
- Nosofsky, R.M.: An exemplar-based random-walk model of speeded categorization and absolute judgment; in Marley, Choice, decision, and measurement: essays in honor of R. Duncan Luce, pp. 347–365 (Erlbaum, Mahwah 1997).
- Palmer, J.; Verghese, P.; Pavel, M.: The psychophysics of visual search. *Vision Res.* 40: 1227–1268 (2000).
- Pisoni, D.B.: Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception Psychophysics* 13: 253–260 (1973).
- Pisoni, D.B.; Tash, J.: Reaction times to comparisons within and across phonetic categories. *Perception Psychophysics* 15: 285–290 (1974).
- Rietveld, A.C.M.; Van Heuven, V.J.: *Algemeine fonetiek* (Uitgeverij Coutinho, Bossum 2001).
- Rothstein, R.A.: Polish; in Comrie, Corbett, *The Slavonic languages*, pp. 686–758 (Routledge, London 1993).
- Rubin, P.; Turvey, M.T.; van Gelder, P.: Initial phonemes are detected faster in spoken words than in non-words. *Perception Psychophysics* 19: 394–398 (1976).
- Savin, H.B.; Bever, T.G.: The non-perceptual reality of the phoneme. *J. verbal Learn. verbal Behav.* 9: 295–302 (1970).
- Schweickert, R.: Information, time, and the structure of mental events: a twenty-five-year review; in Meyer, Kornblum, *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, pp. 535–566 (MIT Press, Cambridge 1993).
- Sebastián-Galles, N.; Cuetos, F.; Carreiras, M.; Martí, M.A.: *LEXESP. Léxico formatizado del español* (Universidad de Barcelona, Barcelona 2000).
- Sharp, D.; Scott, S.K.; Cutler, A.; Wise, R.J.S.: Lexical retrieval constrained by sound structure: the role of the left inferior frontal gyrus. *Brain Lang.* 92: 309–319 (2005).
- Stevens, K.N.: Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. acoust. Soc. Am.* 111: 1872 (2002).
- Stevens, K.N.; Blumstein, S.E.: Invariant cues for place of articulation in stop consonants. *J. acoust. Soc. Am.* 64: 1358–1368 (1978).
- Theeuwes, J.: Perceptual selectivity for color and form. *Perception Psychophysics* 51: 599–606 (1992).
- Van Ooijen, B.: *The processing of vowels and consonants* (Holland Academic Graphics, The Hague 1994).
- Venables, W.N.; Ripley, B.D.: *Modern applied statistics with S-Plus* (Springer, New York 2002).
- Vitevitch, M.S.; Luce, P.A.: Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Memory Lang.* 40: 374–408 (1999).
- Wagner, A.; Ernestus, M.; Cutler, A.: Formant transitions in fricative identification: the role of native fricative inventory. *J. acoust. Soc. Am.* 120: 2267 (2006).
- Warner, N.; Smits, R.; McQueen, J.M.; Cutler, A.: Phonological and statistical effects on timing of speech perception: insights from a database of Dutch diphone perception. *Speech Commun.* 46: 53–72 (2005).
- Whalen, D.H.: Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. *J. acoust. Soc. Am.* 69: 275–282 (1981).
- Zygis, M.; Hamann, S.: Perceptual and acoustic cues of Polish coronal fricatives. *Proc. 15th ICPhS*, pp. 395–398 (2003).