Cross-language similarities and differences in the uptake of place information

Anita Wagner[a)]

Max Planck Institute for Psycholinguistics, Nijmegen, 6500 AH, The Netherlands

Cross-language differences in the use of coarticulatory cues for the identification of fricatives have been demonstrated in a phoneme detection task: Listeners with perceptually similar fricative pairs in their native phoneme inventories (English, Polish, Spanish) relied more on cues from vowels than listeners with perceptually more distinct fricative contrasts (Dutch and German). The present gating study further investigated these cross-language differences and addressed three questions. (1) Are there cross-language differences in informativeness of parts of the speech signal regarding place of articulation for fricative identification? (2) Are such cross-language differences fricative-specific, or do they extend to the perception of place of articulation for plosives? (3) Is such language-specific uptake of information based on cues preceding or following the consonantal constriction? Dutch, Italian, Polish, and Spanish listeners identified fricatives and plosives in gated CV and VC syllables. The results showed cross-language differences in the informativeness of coarticulatory cues for fricative identification: Spanish and Polish listeners extracted place of articulation information from shorter portions of VC syllables. No language-specific differences were found for plosives, suggesting that greater reliance on coarticulatory cues did not generalize to other phoneme types. The language-specific differences for fricatives were based on coarticulatory cues into the consonant. © 2013 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4802904]

I. INTRODUCTION

To identify phonemes, listeners integrate acoustic information that is spread across the utterance. Which acoustic information distinguishes speech sounds for listeners depends partly on the contrasts incorporated in their native phoneme inventories (e.g., Beddor *et al.*, 2002; Manuel, 1987). Speech segments that seem similar across languages can be distinguished by different sources of information (e.g., Recasens, 1987; Li *et al.*, 2011), and listeners rely on acoustic information in language-specific ways (see, e.g., Crowther and Mann, 1992). Also, the same acoustic signal may be processed differently in two languages (e.g., Iverson *et al.*, 2003), because the informativeness of parts of the speech signal depends partly on the native phonology (e.g., Broersma, 2005; Crowther and Mann, 1992). This paper reports a gating experiment examining the informativeness of coarticulatory cues to place of articulation across four languages with very different fricative inventories.

Formant transitions, as the most intrinsic coarticulatory effect, have been acknowledged as cues to place of articulation across decades of research (Liberman *et al.*, 1954; Delattre *et al.*, 1955; Sussman *et al.*, 1998). In the case of plosives, formant transitions and release burst both provide place of articulation information (e.g., Stevens and Blumenstein, 1978). Interestingly, the relevance of transitions and bursts appears reciprocal: the decrease of the

perceptual weight of one leads to an increase in the perceptual weight of the other (Dorman *et al.*, 1977).

One might assume that the same would be true for fricatives. Acoustic studies, however, indicate a minor role for transitional cues for place distinction in fricatives (Jongman, 1998; Jongman *et al.*, 2000). Perceptual studies show that transitional information does affect listeners' identifications, but more so for some fricatives than for others. Harris (1958) had American English listeners identify natural tokens of syllables combining the fricatives /f v θ ð s z ʃ ʒ/ with /i e o u/. Some syllables were cross-spliced such that the vocalic and frication noise information was mismatched. Listeners accurately categorized the sibilants, disregarding information in the vowel. The fricatives /f/ and /θ/, however, were often confused, and their identification improved when they were presented with coherent vocalic information. [Heinz and Stevens (1961) obtained similar results with synthesized fricatives.] Furthermore, Whalen (1981) investigated the effect of vowel context on fricative identification by combining noises from a synthetic continuum between /s/ and /ʃ/ with natural vowels and formant transitions. He concluded that the contribution of formant transitions to fricative identification may indeed be similar to the role they play for plosives: the perceptual weight of transitions may increase when the fricative noise is ambiguous.

Shifts in the perception of a fricative's noise appear further to depend not only on the transitions to an adjacent vowel, but also on the vowel itself. Mann and Repp (1980) and Whalen (1981) showed that listeners' identification of a syllable consisting of a synthetic noise ambiguous between /s/ and /ʃ/ combined with natural vowels is affected by both the transitions and the vowel: the ambiguous segment was more often identified as /s/ when followed by /u/ than by /a/.

[a)]Author to whom correspondence should be addressed. Present address: Dept. of Speech Hearing and Phonetic Sciences, University College London, London, UK. Electronic mail: a.wagner@ucl.ac.uk

Listeners thus adjusted their perceptual evaluation of the frication depending on the rounding of an adjacent vowel. Additionally, Hedrick and Ohde (1993) investigated listeners' perception of fricative place of articulation as a function of frication duration, vowel quality, formant transitions, and frication amplitude relative to vowel amplitude. The best identifications resulted from amplitude comparisons between frication and the onset of the vowel; the information gain provided by this overrode coarticulatory effects of vowel quality and formant transitions.

Since a number of coarticulatory cues can carry information about place of articulation, listeners with different native backgrounds may differ in terms of the cues on which they rely. Wagner et al. (2006) discovered cross-language differences in the reliance on coarticulatory cues for fricatives. Dutch, English, German, Polish, and Spanish listeners performed a phoneme monitoring task with natural, cross-spliced, or identity-spliced materials, in which the fricative targets were surrounded by vowels containing either coherent or mismatching information. For example, the /s/-noise in the nonsense word *tikusa* was either spliced in from another token of *tikusa* (identity-spliced), or it replaced the /f/-noise as recorded in *tikufa* (cross-spliced). While conflicting cues did not impair fricative identification by Dutch and German listeners, they had a strong adverse effect on English, Polish, and Spanish listeners. Moreover, English and Spanish listeners were particularly hindered in identification of /f/, while Polish listeners made more errors identifying /s/.

Wagner et al. (2006) hypothesized that these language-specific differences arise from the fricative repertoires of the languages tested. Dutch and German fricative contrasts are spectrally very distinct, but English and Spanish have the spectrally similar /f/ and /θ/, while Polish distinguishes four palatal sibilants. In learning to distinguish contrasts that are perceptually quite similar, Spanish, English and Polish listeners have learned that the vocalic portion adjacent to the frication is helpful and informative.

The study by Wagner et al. (2006) suggested that listeners differ in how they extract information specifying place of articulation for fricatives. Their study also showed that listeners may make no use of systematic acoustic variation in the signal. As a consequence, listeners of different languages may differ in the amount of information they have about a speech sound at given time points as the signal unfolds.

The study by Wagner et al. (2006) leaves open the question of the relative value of cues preceding or following the frication, since their listeners heard conflicting vocalic portions surrounding the frication. For plosives, studies exploring the perceptual effect of formant transitions tend to focus on the transitions following the burst than transitions preceding the closure (e.g., Stevens and Blumenstein, 1978; Liberman et al., 1954; Delattre et al., 1955; Sussman et al., 1998), and several studies report that post-consonantal transitions outweigh the pre-consonantal transitions (e.g., Dorman et al., 1975; Fujimura et al., 1978). Other studies, however, report a rather moderate predominance of post-consonantal transitions (Tartter et al., 1983; Repp, 1978), and stress the interaction of pre-consonantal and post-consonantal transitions with other cues like closure duration.

For fricatives, Mann and Soli (1991) compared the contribution of formant transitions across fricative-vowel (FV) and vowel-fricative (VF) syllables. Transitions showed a greater effect on listeners' identification in FV than in VF syllables. However, when listeners heard FV and VF formant transitions played in reversed order, their identification was determined by the order of presentation, and not by an intrinsic difference between transition types. Nittrouer et al. (2000), using the same paradigm, likewise showed that adult listeners made more use of post-frication than of pre-frication formant transitions. In this study a modest effect of order of presentation was found. VF transitions in their materials, however, appeared to contain less information about the fricative than FV transitions.

The present study tested listeners' use of coarticulatory cues as a function of the number of places of articulation for fricatives contrasted in their native phoneme inventory. Three questions are posed, all arising from the previously observed differences in the reliance on coarticulatory cues for fricatives. (1) Are there cross-language differences in informativeness of parts of the speech signal regarding place of articulation for fricative identification? (2) Are such cross-language differences fricative-specific, or do they extend to the perception of place of articulation for plosives? (3) Is such language-specific uptake of information based on cues preceding or following the consonantal constriction? Listeners from four different native backgrounds were compared in a gating experiment.

Gating is a proven task for studying how continuously evolving acoustic information is perceptually evaluated at given time points (Grosjean 1980; Smits, 2000; Smits et al., 2003, Warner et al., 2005). In gating, listeners hear truncated portions of speech signals (for an overview see Grosjean, 1996), the procedure thus allows an assessment of the acoustic information available in different portions of the signal. Gating studies show that listeners base their identification of speech segments on information spread across the utterance. The gating technique can be used in two different ways: forward gating and backward gating. In forward gating listeners are presented with parts of the signal preceding the truncation point, while in backward gating listeners hear portions of the signal following the truncation point. Studies using backward gating have shown the relevance of cues in the vowel portions following, and studies using forward gating the relevance of acoustic events preceding the constriction of a consonant (e.g., Smits, 2000; Smits et al., 2003; Warren and Marslen-Wilson, 1987). The gating paradigm thus makes it possible to address the question of cross-language differences in informativeness of coarticulatory cues preceding and following a consonant.

II. EXPERIMENT

A. Language comparison

We compared Dutch, Polish, Spanish, and Italian; the first three were also among the languages tested by Wagner et al. (2006). In that study, a mismatch between vowel and frication impaired fricative identification for Polish and Spanish listeners, whose native fricative repertoires contain perceptually similar fricative pairs, but not for

Dutch or German listeners, who have only perceptually robust fricative contrasts. Italian allows us to test the generalization of the Dutch and German pattern to another language with disparate contrasts. Although the languages differ in their fricative inventories, they all have three places of articulation for plosives: labial, alveolar, and velar.

Dutch distinguishes fricatives at four places of articulation, labiodental /f v/, alveolar /s z/, velar /x/, and glottal /h/. The Italian fricative inventory contains five spectrally distinct fricatives at three places of articulation: labiodental /f v/, alveolar /s z/, and palatal /ʃ/. In acoustic terms Italian sibilants can be described as distinct because there is no overlap in spectral distribution of frication peaks: /s/ peaks between 4–7 kHz, /ʃ/ between 2–4 kHz (Sorianello, 2002). Compared to these two languages, the other two inventories are more complex. Polish has 11 fricatives at six places of articulation, among them four coronal places of articulation for sibilants: alveolar /s z/, postalveolar /ʃʲ ʒʲ/, retroflex /ʂ ʐ/, and alveolo-palatal /ɕ ʑ/. In acoustic terms, energy peaks of the fricative /s/ are typically in the frequency range between 3 and 7 kHz, of the post-alveolar /ʃʲ/ between 1.5 and 5 kHz, of the alveolopalatal /ɕ/ between 2 and 6 kHz, and of the retroflex /ʂ/ around 1–4 kHz (Jassem, 1968; Lipski, 2006; Lisker, 2001). The coronal Polish fricatives thus show an overlap of energy peaks across their noise spectra. Spanish contrasts fricatives at four places of articulation. Next to alveolar /s/ and velar /x/ Spanish also has the spectrally similar labiodental /f/ and dental /θ/. These two are very similar because their spectra show a relatively flat distribution of energy across a wide range of frequencies from 2 to 10 kHz, with no distinct energy peaks (see Jongman et al., 2000 for an acoustic analysis of corresponding English categories).

Among these languages there is less variation regarding the number of plosive places of articulation, but the exact realization of these phonemes will be language-specific nonetheless. Typically Spanish, Italian, and Dutch plosives are described as unaspirated, and Polish plosives as slightly aspirated, but, in the final position Polish and Dutch plosives are produced with aspiration. An aspiration following the burst may result in a slower release, and a shorter closure duration (Cho and Ladefoged, 1999). This acoustic information may be weighted in language-specific ways.

Importantly, many factors contribute to language-specific realizations of phonemes. Among these factors are the rhythmic patterns of the language, language-specific phonotactics, or assimilation rules. These factors can restrict the occurrence of some phonemes in certain positions. For instance, word final consonants are always devoiced in Polish and Dutch, and Spanish allows only few consonants in word final position. Also, dialectal variation creates variability in phonetic realization even within a language. All these factors contribute to the processing of speech, but possibly not in the experimental paradigm used here. The present study takes advantage of the fact that listeners, when presented with a non-native realization of a phoneme, inevitably assimilate the heard sound to their most similar native category. Listeners from different native backgrounds will be presented with a phonetic realization that is native or closer to native for some listeners, and less so for others. As

shown in previous studies (Wagner et al., 2006; Wagner and Ernestus, 2008), in such an approach listeners rely on language-specific cues as well as on acoustic sources that are generally inherent to sound categories [in analogy to the acoustic landmarks of Stevens (2002)].

On the basis of the results of Wagner et al. (2006), we first predict cross-language differences in the informativeness of coarticulatory cues concerning place of articulation in fricative identification: listeners with several similar places of articulation (Polish, Spanish) will gain more information from other (vocalic) portions of the speech signal than listeners with more robust fricative categories (Dutch, Italian). Second, because these languages do not differ in their places of plosive articulation, we can test whether reliance on coarticulatory cues is specific to fricatives or whether it indicates language-specific differences in reliance on transitional cues in general; only if sensitivity to coarticulatory cues generalizes across phoneme types will differences be observed with plosives. In such a case, the differences would be expected to resemble those of fricatives.

B. Method

1. Materials

The plosives [p t k] and the fricatives [f s] were combined with the vowels [a i u] to create 15 CV and 15 VC syllables, e.g., *af*, *ip*, *ut* for forward-gated, *pa*, *su*, *ki* for backward-gated materials. The materials were recorded by a phonetically trained Dutch speaker in a sound-attenuated room directly to a computer and down-sampled to 22.05 kHz (16-bit resolution). The speaker was fluent in English but spent most of his life in the Netherlands. He was instructed to record the stimuli as if these were Dutch words. The speaker produced fricatives with spectral peak values between 3 and 7 kHz for /s/, and a flat energy distribution between 2 and 10 kHz for /f/. The final plosives were produced with an aspiration ranging from 180 ms for /t/ to 100 ms for /p/.

a. Gating procedure. The gated materials were constructed with PRAAT software. The points of truncation were defined visually on the basis of the waveform and a wideband spectrogram. Forward gating was used for VC syllables and backward gating for CV syllables. Figure 1 illustrates the placement of truncation points for the materials, and displays the number of gates for both fricative and plosive targets.

First, the offset (for VC syllables) and onset (for CV syllables) of voicing in the vowel were defined as onset or offset of the consonantal constriction. Second, truncation points in the vowel portion adjacent to the constriction were added in 20-ms increments. The shortest gates contained a portion of the vowel starting (for CV syllables) or ending (for VC syllables) 40 ms after (before) the constriction. In a third step the consonants were gated. In this step differences between the acoustic properties of fricatives versus plosives caused differences in the exact duration of gates. For fricatives, the longer and continuous frication was gated in two 20-ms increments, and the longest gate contained the entire frication. The shorter and abrupt plosives allowed only four gates in CV syllables: one consonantal gate containing the release burst (the silent interval of the closure was not included, and syllable initial

plosives were produced without aspiration). In VC syllables the release burst was defined as the first truncation point, and the longest gate included also a period of aspiration. As a consequence, for fricatives the longest gates contained a frication of at least 120 ms, while the longest gates for plosives in CV syllables included only the short release burst, and in VC syllables the silent interval of the closure, the release burst and a short period of aspiration. This procedure resulted in six gates for fricatives and plosives in VC syllables, and in six gates for fricatives, but only four gates for plosives in CV syllables. In total, this resulted in 90 truncations (5 targets × 3 vowels × 6 gates) in VC sequences and 72 truncations (3 plosive targets × 3 vowels × 4 gates, and 2 fricative targets × 3 vowels × 6 gates) in CV sequences. Table I lists the average duration of the stimulus per gate, and the duration of the vowel- and consonant-portion contained.

Previous gating experiments have shown that truncating the signal can add clicks or external noises to the gated segments (Pols and Schouten, 1978). To avoid this, and to minimize a previously observed response bias for labials (Smits *et al.*, 1996) a 500 Hz square wave replaced the deleted parts of the signal, and a linear ramp within a window of 5 ms was applied at the truncation points and the square wave. Ten

practice syllables with [d n m] as targets and with the vowels [e u] were recorded by the same speaker. These syllables were gated following the steps described above, where the beginning or end of the nasal murmur defined the onset and offset of the consonant. The practice syllables served to familiarize the participants with the gating task, and with the somehow unnatural sound of the truncated stimuli.

2. Procedure

The experiment was carried out in four different locations. At each, listeners sat in sound attenuated booths, and were presented with the materials over headphones. Participants were instructed in their native language. The task was to label the consonant preceding (CV syllables) or following (VC syllables) the vowel, as one of /f s k p t/. The responses were given onto a keyboard, where five adjacent keys were marked as the response options, and labeled with the letters f, s, k, p, and t. The labels were the same for all participants. At each location, the experimenter explained to the participants, in their native language, which native sounds corresponded to which response category by giving them examples of monosyllabic words containing these
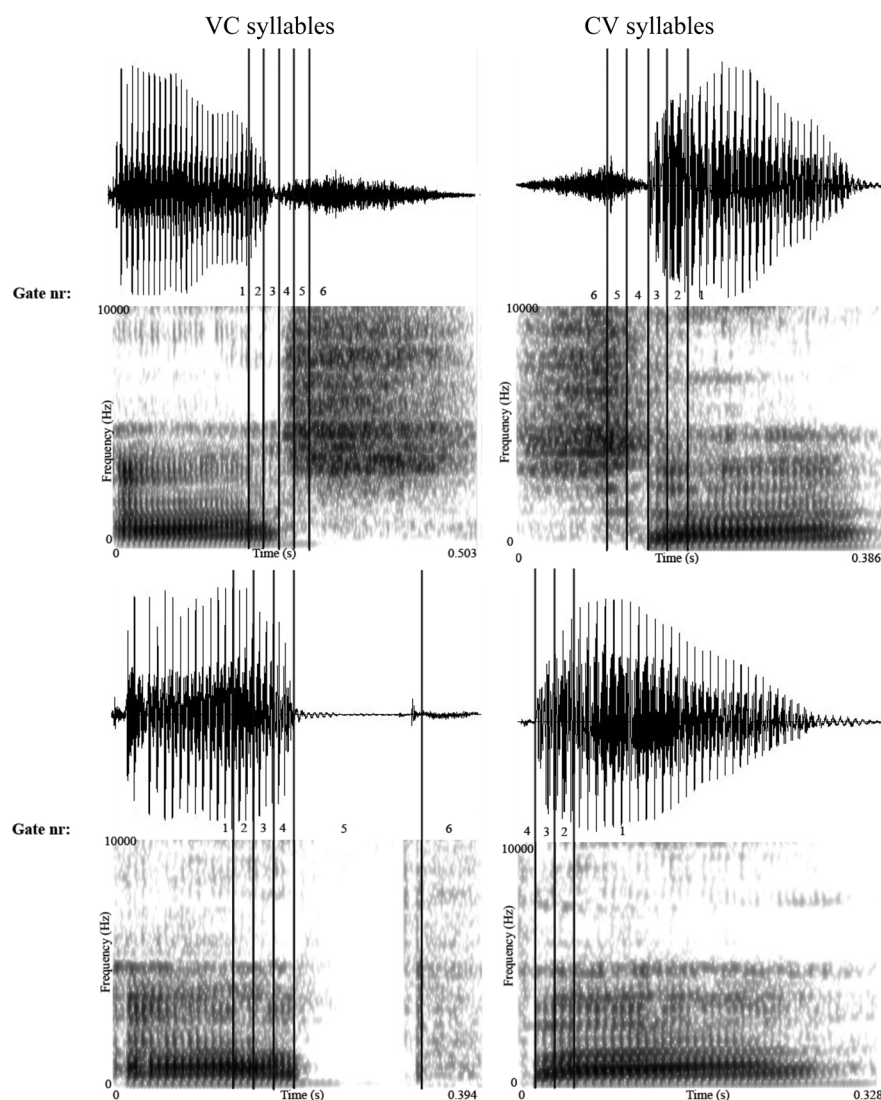

FIG. 1. The placement and the number of gates for fricative and plosive targets in backward-gated (CV) and forward-gated (VC) syllables. Displayed are waveforms and spectrograms of the syllables /as/, /sa/, /ap/, and /pa/.

TABLE I. Duration of the stimuli, and of the contained consonant and vowel portions in each gate in CV- and VC-syllables for the plosive- and fricative-targets. The numbers are durations in ms of the plosives (k, p, t) and fricatives (f, s) averaged across the three context vowels.

		VC				CV		
Fricatives	gate	stimulus	vowel	fricative	gate	stimulus	fricative	vowel
	1	100	100	—	1	210	—	210
	2	120	120	—	2	230	—	230
	3	140	140	—	3	250	—	250
	4	160	140	20	4	270	20	250
	5	180	140	40	5	290	40	250
	6	420	140	280	6	400	150	250
Plosives	gate	stimulus	vowel	plosive	gate	stimulus	plosive	vowel
	1	110	110	—	1	195	—	195
	2	130	130	—	2	215	—	215
	3	150	150	—	3	235	—	235
	4	170	170	—	4	270	35	235
	5	270	170	100	5	—	—	—
	6	400	170	230	6	—	—	—

phonemes. A selection of native words for each target phoneme was used to exemplify that the task was to map response letters to native sounds and not to orthographic symbols. Ten practice trials were used to familiarize the participants with the experimental task.

From the onset of each item listeners had ten seconds to respond, and were instructed to guess in cases where they were very uncertain. The presentation of the next item started 1000 ms after a response was given, on a keyboard by pressing keys labeled with one of the five response options. The stimuli were blocked into forward-gated and backward-gated materials. Within these two blocks the materials were presented in random order. For each participant a different pseudo-random order of presentation was chosen. Between the two blocks was a pause, the duration of which was controlled by the participants themselves. The order of presentation of the blocks was counterbalanced over the participants.

3. Participants

Fifteen native Dutch speakers from the participant pool of the Max-Planck-Institute for Psycholinguistics in Nijmegen, nine native Italian students at the University of Trieste, 15 Polish native speakers, students at the Universitet Slaeski in Katowice, and 18 Spanish native speakers, students at the Universitat de Barcelona, participated in this experiment. None of them reported any speech or hearing disorders. All listeners spoke no other language than their native up to an age of, at least, seven. Possible other languages spoken by the participants, like Catalan for most of the Spanish participants or English for most of the Dutch participants, were acquired only in school. Participants received a small payment or were rewarded with course credits.

4. Analysis of results

The data will first be presented in terms of percentages of correct responses. Correct responses, however, are limited in their explanatory power, because they only show the cases in which listeners recognized both place and manner correctly. More insight will be derived from listeners' confusion patterns; with materials of very short duration listeners' guesses reflect the acoustic information which is extracted even though the phoneme is not unambiguously identified. As the focus of this study is the gain of information for place of articulation, an analysis of transmitted information (TI) as a function of gate was conducted. TI is a measure of covariance of responses and stimuli, and gives insight about information reducing listeners' uncertainty about a response as a function of the stimulus (Shannon, 1948; Miller and Nicely, 1955; Jongman, 1989; Smits, 2000). This measure is used for categorical judgments, per phonological feature, and is calculated from the entire set of responses. One advantage of this measure is that it is bias-free, with values next to 0 if no information about the response was transmitted by the stimulus, and values approaching 1 if the stimulus transmitted all the information needed for a correct response. TI is calculated on the basis of the following formula:

$$\text{TI}_{(S,R)} = \Sigma_s \Sigma_r p(s,r) * \log p(s,r)/p(s)p(r),$$

where $p(s,r)$ is the probability of the joint occurrence of the response r with the stimulus s, $p(r)$ is the marginal probability of the response and $p(s)$ is the marginal probability of the stimulus. The maximum of TI within a set of responses is expressed as the entropy of the stimulus, and described as

$$H(s) = -\Sigma_s p(s) * \log p(s).$$

Whereas TI expresses the transmission from stimulus to response in bits, the relative information transmitted is given by

$$\text{TI rel}_{(S,R)} = T_{(S,R)}/H(s).$$

C. Results and discussion

1. Correct responses

Responses that did not occur within ten seconds from the onset of the item (1,7% in CV syllables, and 4,2% in VC syllables) were excluded from the analyses, as were the data of one Dutch participant who missed more than 40% of all items.

Table II lists the percentages of correct responses for fricatives and plosives for the four languages in CV syllables. Table III presents the same for VC syllables. The probability of a correct response was analyzed in two logistic models with mixed effects (Venables and Ripley, 2002; Baayen, 2008), with Language (Dutch, Italian, Polish, Spanish), gate (gate 1–6, or gate 1–4 for plosives in CV syllables), and Phoneme Type (fricative or plosive) as independent factors and Participant as random factor, separately for CV- and VC-syllables. Two analyses of variance (ANOVAs) were then applied to the models to test for statistical effects of the independent factors on the dependent variable correct responses. In mixed-effect models, degrees of freedom cannot be considered as the sample size minus the number of free parameters. More precisely, it is disputable how many parameters are added to a model by including one random factor. It is thus a debated topic how to determine

TABLE II. Percentage of correct responses for fricatives and plosives in CV syllables for the four listeners groups.

	Fricatives				Plosives			
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	19	21	11	8	1	16	24	24	28
2	12	18	15	12	2	20	32	24	31
3	11	16	21	12	3	47	40	27	46
4	2	7	10	6	4	98	90	90	92
5	11	17	12	12					
6	100	98	100	99					

significance levels in mixed-effect models [for a discussion see Baayen *et al*. (2008)]. In the present study the denominator degrees of freedom were 3961 for the CV model, and 4824 for the VC model. To keep the statistics conservative, the *F* values were calculated and then a lower bound was set to the degrees of freedom. Following Reubold *et al*. (2010), an α-level of 0.001 for 60 degrees of freedom was chosen, which means that *F* values > 11.97 are considered significant at $p < 0.001$.

The main effect of Language did not emerge as significant ($F = 0.6354$), and accordingly Fig. 2 displays the correct identifications as a function of gate averaged across all listeners. Percentages of correct identifications as a function of gate, separately for fricatives and plosives in CV syllables are displayed in the top panel of Fig. 2, and the bottom panel displays the same for VC syllables.

In CV syllables, the significant effects were Gate ($F = 234.10$), Phoneme Type ($F = 387.37$), and the interaction between Gate and Phoneme Type ($F = 15.65$). In VC syllables the significant effects were also Gate ($F = 612.06$), Phoneme Type ($F = 258.62$), and the interaction between Gate and Phoneme Type ($F = 150.19$).

The effect of Phoneme Type and the interaction of Gate and Phoneme Type indicate a difference in how listeners identified plosives versus fricatives. However, these effects may also reflect listeners' bias towards plosives, reported in previous gating studies (e.g., Klaassen-Don, 1983): fricatives are identified less accurately because the frication duration needs to exceed a certain length in order not to be perceived as a release burst. Also, correct responses reflect only the cases where listeners correctly identified both the manner and the place of articulation. Theses results cannot tell apart whether these responses were based on information about place of articulation or on a bias towards plosives. Therefore, listeners' responses were

TABLE III. Percentage of correct responses for fricatives and plosives in VC syllables, for the four listener groups.

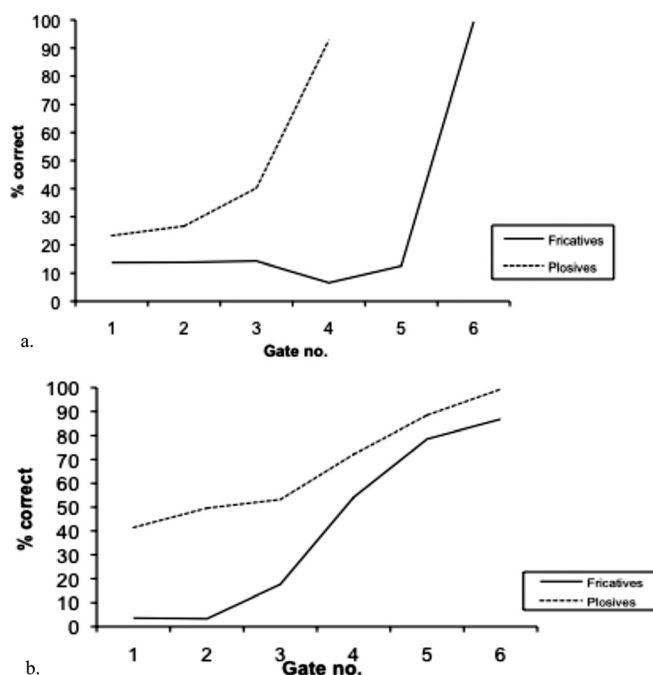|	Fricatives				Plosives			
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	5	6	3	2	1	45	43	37	41
2	6	4	3	1	2	48	45	44	58
3	25	28	14	9	3	55	48	52	54
4	52	55	61	46	4	73	72	69	74
5	86	87	85	63	5	94	90	88	84
6	100	100	96	95	6	100	100	99	98

FIG. 2. Percentage of correct responses averaged across all listener groups, for fricatives and plosives respectively as a function of gate in CV-syllables (a) and VC-syllables (b).

further analyzed in terms of TI for manner and place of articulation. For the analysis of TI, listeners' responses were pooled into confusion matrices (for the matrices see Wagner, 2008, pp. 115–118), from which matrices were extracted for manner (irrespective of place of articulation), and for place of articulation, separately for fricatives and plosives. For place of articulation this means that /p/ and /f/ were treated as one place of articulation, and /t/ and /s/ as another.

2. Transmitted information

a. Manner of articulation analysis. Table IV displays the percentages of correct responses for manner (independent of place) in CV and VC syllables. Table V displays correct responses for place (independent of manner) for CV and VC syllables, and serves to illustrate general patterns in the responses, but note that correct responses also reflect a bias towards plosives. Figure 3, left panel, displays the percentages of TI for manner of articulation as a function of gate for the four languages in CV syllables, and the right panel displays the same for VC syllables. To investigate which gates caused a significant increase in TI pair-wise Kruskal-Wallis tests were conducted across two consecutive gates, for each language separately. For CV syllables, only the first four gates were included in the analysis of manner of articulation since plosive targets were presented only in four gates in CV syllables. Each comparison was evaluated at a corrected α level of 0.004 (0.05 divided by 12) in CV syllables, and of 0.002 (0.05 divided by 20) in VC syllables.

As displayed in Fig. 3(a), in CV syllables listeners extracted little information about manner of articulation. For all four languages no significant effect of gate was found. Even though Table IV displays correct identifications of manner of articulation of above 50% in the first four gates,

TABLE IV. Percentage of correct identification in terms of manner of articulation (independent of place of articulation) for the four listener groups in CV and VC syllables.

	CV syllables				VC syllables			
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	55	59	54	53	1	61	65	59	60
2	49	61	57	58	2	61	57	55	58
3	58	59	57	56	3	65	65	63	62
4	60	63	62	60	4	81	84	82	80
					5	97	99	97	86
					6	100	100	98	98

the analysis of TI reveals that these identifications in fact reflect a bias towards plosives.

In VC syllables significant increases in TI emerged between gates 3 and 4 for all the languages (Kruskal-Wallis tests, Dutch: $\chi^2 = 11.3$; Italian: $\chi^2 = 13.33$; Polish: $\chi^2 = 16.82$; Spanish: $\chi^2 = 19.55$; df $= 1$, $p < 0.001$) and between gates 4 and 5 for Dutch, Italian, and Polish (Kruskal-Wallis tests, Dutch: $\chi^2 = 18.16$; Italian: $\chi^2 = 9.8$; Polish: $\chi^2 = 16.88$; df $= 1$, $p < 0.002$) but not for Spanish. A t-test showed that the difference between Spanish on the one hand, and Dutch, Polish and Spanish on the other reached significance [t $(71) = 3.00$, $p < 0.004$, effect size: $d = 0.60$].

b. Place of articulation analysis: Fricatives. Figures 4(a) and 4(b) show the percentage of TI for place of articulation as a function of gate and language for CV and VC syllables, respectively. Again, pair-wise Kruskal-Wallis tests were conducted across two consecutive gates, for each language separately. Each comparison was evaluated at a corrected α level of 0.002 (0.05 divided by 20 comparisons).

For CV syllables a significant effect of Gate for all four languages emerged between gates 3 and 4 (Kruskal-Wallis tests, Dutch: $\chi^2 = 13.47$; Italian: $\chi^2 = 12.24$; Polish: $\chi^2 = 15.35$; Spanish: $\chi^2 = 14.51$; df $= 1$, $p < 0.002$), after the presentation of the 20 ms of frication closest to the vowel, and between gates 5 and 6 (Kruskal-Wallis tests, Dutch: $\chi^2 = 24$; Italian: $\chi^2 = 15.68$; Polish: $\chi^2 = 23.84$; Spanish: $\chi^2 = 30.93$; df $= 1$, $p < 0.001$), after the presentation of the entire frication noise. In VC syllables, a difference was found between Spanish and Polish listeners on the one hand, and Dutch and Italian listeners on the other. An effect of gate for Spanish and Polish listeners emerged only between gates 3 and 4 (Kruskal-Wallis tests, Polish: $\chi^2 = 12.44$; Spanish: $\chi^2 = 10.73$; df $= 1$, $p < 0.001$),

TABLE V. Percentage of correct identification in terms of place of articulation (independent of manner of articulation) for the four listener groups in CV and VC syllables.

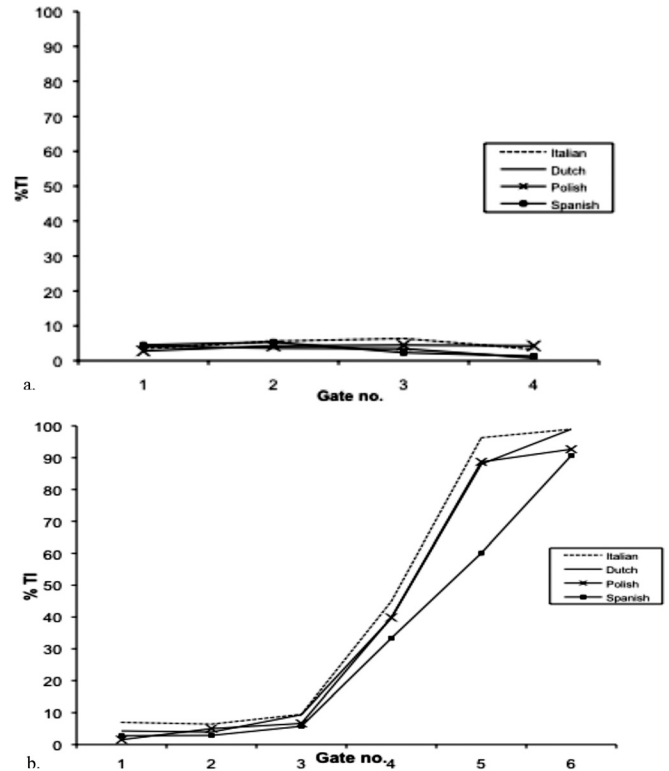|	CV syllables				VC syllables			
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	32	40	32	38	1	49	45	43	46
2	39	40	38	44	2	51	51	51	58
3	54	49	47	60	3	59	55	54	57
4	92	91	93	94	4	74	73	78	77
5	89	83	91	90	5	92	90	89	83
6	100	98	100	100	6	100	100	99	97

FIG. 3. Percentage of TI for manner of articulation as a function of gate per language group in CV (a) and in VC (b) syllables.

after the first 20 ms of frication, whereas a significant increase in TI for Dutch and Italian listeners emerged between gates 4 and 5 (Kruskal-Wallis tests, Dutch: $\chi^2 = 15.23$; Italian: $\chi^2 = 9.88$; df $= 1$, $p < 0.002$), thus after the presentation of 40 ms of frication. The difference between Spanish and Polish, on the one hand, and Dutch and Italian on the other, was significant [t $(107) = 2.56$, $p < 0.02$, effect size: $d = 0.47$].
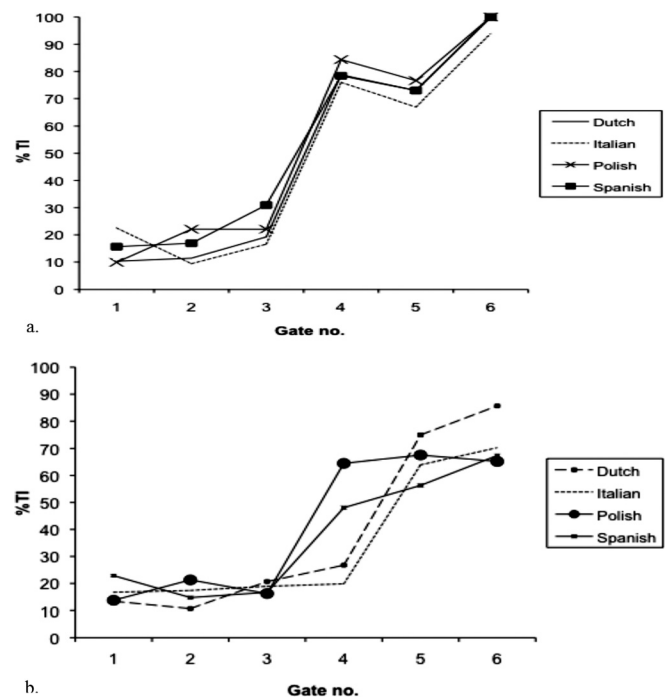

FIG. 4. Percentage of TI for place of articulation as a function of gate per language group for fricative targets in CV (a) and in VC (b) syllables.

Anita Wagner: Language-specific informativeness of cues

TABLE VI. Percentage of correct identifications of place of articulation for fricative targets in VC syllables split by the two targets.

	Dutch		Italian		Polish		Spanish	
Gate	f	s	f	s	f	s	f	s
1	83	26	58	38	58	38	61	43
2	79	31	65	41	78	38	67	48
3	74	50	78	46	62	44	61	56
4	85	55	78	59	91	82	85	76
5	100	81	100	78	98	82	87	78
6	100	100	100	100	96	100	91	100

The analysis of TI for place of articulation captures TI for place of articulation for both /s/ and /f/. The number of observations does not allow splitting the analysis into the two targets. Table VI, however, displays the percentage of correct identifications of place of articulation for the four language groups and the six gates divided into the two fricative targets, and suggests that the increase in TI for Polish and Spanish listeners is carried by more accurate identification of both /f/ and /s/ in gate 4. The identification of place of articulation for Dutch and Italian listeners improves in gate 5, and in particular for /s/.

c. Place of articulation analysis: Plosives. Figure 5(a) presents the TI per language group as a function of gate for plosives in CV syllables, and Fig. 5(b) displays the same for VC syllables. TI for plosives was analyzed in the same way as for fricatives, separately for the languages, and across consecutive gates. The comparisons were evaluated at a corrected α-level of 0.002 (0.05 divided by 20) in VC syllables, and of 0.004 in CV syllables (0.05 divided by 12).

In CV syllables all languages show a significant increase in TI only in gate 4 (Kruskal-Wallis tests, Dutch: $\chi^2 = 21.88$; Italian: $\chi^2 = 12.21$; Polish: $\chi^2 = 19.07$; Spanish: $\chi^2 = 26.36$; df = 1, $p < 0.004$), after the presentation of the burst noise. In VC syllables, all languages show a significant increase in TI between gates 3 and 4 (Kruskal-Wallis tests, Dutch: $\chi^2 = 10.84$; Italian: $\chi^2 = 12.48$; Polish: $\chi^2 = 9.65$; Spanish: $\chi^2 = 9.49$; df = 1, $p < 0.002$), and 5 and 6 (Kruskal-Wallis tests, Dutch: $\chi^2 = 23.93$; Italian: $\chi^2 = 15.3$; Polish: $\chi^2 = 25.89$; Spanish: $\chi^2 = 19.93$; df = 1, $p < 0.002$). Thus the amount of information rises gradually as a function of the presentation of the vocalic portion preceding the closure, and the burst.

In summary, the present study showed some common patterns across the languages tested. Regarding the manner of articulation, in CV syllables all listeners hardly gained any information. A reason for this might be that the presence of silence during closure versus the presence of frication serves as a major cue to manner of articulation, and it is impossible to determine the closure onset in CV syllables in the present stimuli. Another reason might be that the gradual amplitude increase in the noise spectrum, as suggested by Jongman (1989), is the relevant cue to manner for fricatives. This slow increase in amplitude is contained in the onset portion of the fricative. In the present study listeners were presented with the signal portion closest to the vowel, and the frication onset was presented only within the entire syllable, thus in the last gate.
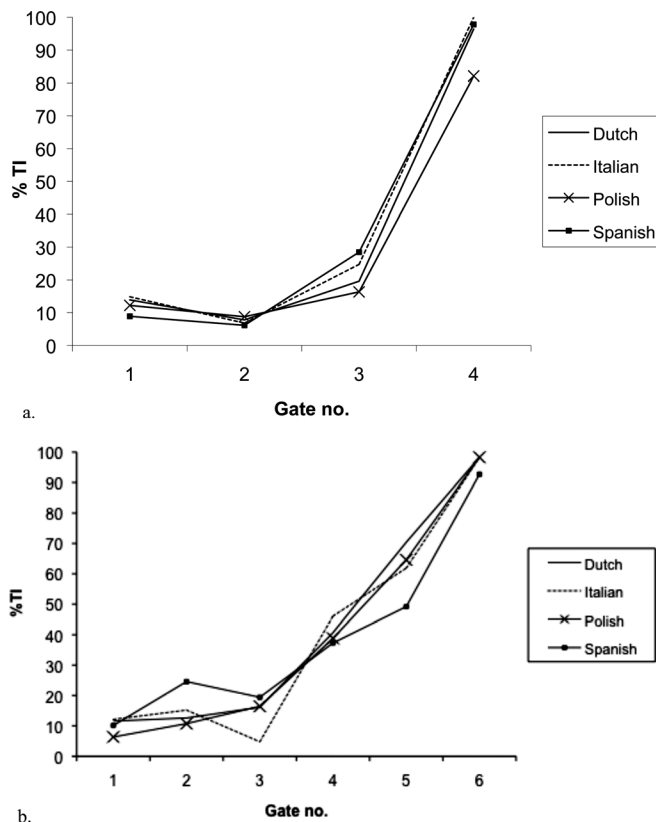

FIG. 5. Percentage of TI for place of articulation as a function of gate per language group for plosive targets in CV (a) and in VC (b) syllables.

The gain of manner information in VC syllables shows, unsurprisingly, that the presence of closure versus frication is the main cue for manner of articulation for all the listener groups tested. A possible explanation for a smaller gain of information for Spanish listeners for the portion containing the release burst versus 40 ms of frication may be that in Spanish plosives in syllable final positions may be realized as fricatives or approximants (Martinez-Celdran et al., 2003). The distinction of manner of articulation thus loses relevance for consonants in syllable final position.

In fricative identification, all listeners extracted information about place of articulation from a portion of the frication and not from parts of the vowel adjacent to the frication. This result is in line with the data from the large-scale diphone gating experiments with Dutch stimuli and participants of Smits et al. (2003) and Warner et al. (2005). These authors showed that the first third of the frication contains substantial information about the place of articulation, and that the information in the vowel adjacent to the frication does not improve identification. Also, Jongman (1989) presented English listeners with fricative noises of durations varying in 10-ms-steps from 20 to 70 ms from the onset of the frication, and found that listeners needed more than 40 ms to correctly identify the place of articulation. In line with this, in the present study listeners' identification of place of articulation in CV syllables improved when a part of the frication (20 ms closest to the vowel) was presented, a slightly longer frication portion (40 ms) did not improve the identification any further, and the crucial place information appeared to be transmitted when the whole frication (120 ms) including the onset was presented. The results of these studies

combined suggest that the crucial information specifying a fricative's place of articulation is contained in the first 80 ms of frication.

Importantly, the present study also found a language-specific pattern for the perception of place for fricatives in VC syllables. Only Spanish and Polish listeners gained place of articulation information from 20 ms of frication following the vowel, presumably because these listeners differentiate more places of articulation in their native language. Dutch and Italian listeners, thus listeners with only spectrally disparate fricative categories, gained this information from a longer portion (40 ms) of frication.

For the voiceless plosives in the present study all listeners used the information in the release burst. In VC syllables all listeners gained place information also from the vocalic portion preceding the constriction. It might seem surprising that listeners gained so little information about place of articulation from the vocalic portion in CV syllables, especially since many studies investigating formant transitions focus on the post-consonantal transitions (e.g., Stevens and Blumenstein, 1978; Liberman et al., 1954; Delattre et al., 1955; Sussman et al., 1998). It is probable, however, that the formant transitions in the present study carried less place information because formant transitions in voiceless plosives are shorter (Benki, 2001). In line with Dorman et al. (1977), in the present study, it was mainly the release burst that carried place information so as to compensate for the smaller perceptual weight of the formant transitions.

III. CONTROL EXPERIMENT

The study by Wagner et al. (2006) found no differences in the reliance on coarticulatory cues between materials produced by a Dutch speaker versus materials by a Spanish speaker. In order to exclude the possibility that in the present gating study listeners were sensitive to language-specific realizations of phonemes, an additional experiment was conducted with materials produced by a Spanish speaker, on a subset of participants from the main experiment.

A. Materials

A phonetically trained Spanish speaker recorded the set of syllables from the main experiment. The speaker was fluent in English and familiar with Dutch, but his dominant language was Spanish, as spoken in the north-western parts of Spain. The speaker was instructed to produce the syllables as if these were Spanish words. Similar to the Dutch speaker, the Spanish speaker produced fricatives with a spectral peak between 4 and 7 kHz for /s/, and a flat distribution of energy without any peaks for /f/. The plosives, however, were produced without aspiration. The materials were processed and gated in the same way as in the main experiment, but the absence of aspiration in the Spanish stimuli allowed for only five gates for plosives in VC syllables.

B. Procedure and participants

The procedure was as in the main experiment. The 15 Polish participants and 18 Spanish participants from the main experiment took part in this control experiment.

C. Results and discussion

Results were analyzed in the same way as in the main experiment. Results from the control experiment were then pooled together with the results for the Spanish and Polish participants from the main experiment (for correct responses and for TI). The probability of a correct response was analyzed in two logistic mixed-effects models with Language (Polish, Spanish), Speaker (Dutch and Spanish), Gate (gate 1 to 6 for fricatives, gate 1 to 4 for plosives in CV syllables, and gate 1 to 5 for plosives in VC syllables), and Phoneme Type (fricative or plosive) as independent factors and participant as random factor, separately for CV- and VC-syllables. Two ANOVAs were then applied to the models.

In CV syllables there was no main effect of Speaker ($F = 5.33$), and no interaction with Speaker emerged significant. As in the main experiment, the significant effects were Gate ($F = 285.11$), Phoneme Type ($F = 1179.74$), the interaction between Gate and Phoneme Type ($F = 122.59$). Also in VC syllables Speaker did not appear as significant ($F = 0.89$). Also here, the significant effects were Gate ($F = 2387.93$), Phoneme Type ($F = 615.67$), and the interaction between Gate and Phoneme Type ($F = 336.19$).

In line with this, the analysis of TI for the Spanish and Polish participants, when the results from both experiments were pooled together, showed no effect of speaker, and no significant interactions with speaker. This shows that, in the present experiment, Polish and Spanish listeners extracted information from the same portions of the acoustic signal irrespective, of whether the stimuli were produced by a Spanish or by a Dutch speaker. Furthermore, these results consolidate the cross-language difference found for fricatives in VC syllables: there was no interaction between Speaker and Gate in the TI analysis.

IV. GENERAL DISCUSSION

The present study was designed to test the following questions: (1) are there cross-language differences in the informativeness of coarticulatory cues to place of articulation for fricatives; (2) are such cross-language differences specific to fricatives; (3) are these differences based on the information preceding or following the consonant constriction? Listeners with Dutch, Italian, Polish, and Spanish as native languages were presented with truncated portions of CV and VC syllables with fricative and plosive targets. No cross-language differences were found in the perception of place for plosives. For fricatives the results show that Spanish and Polish listeners extract place information from a relatively short portion of frication in VC syllables, while Dutch and Italian listeners require a longer portion of the frication to gain a similar amount of place information.

The language-specific difference in the uptake of place information for fricatives confirms and further specifies the findings by Wagner et al. (2006) with a different experimental paradigm: Spanish and Polish listeners gain information about place of articulation for fricatives on the basis of information from the preceding vowel and a shorter portion of the frication. This result supports the conclusion that listeners with acoustically similar native fricative pairs rely on more sources of information in order to identify all contrasts accurately. Moreover,

this study specifies that these listeners rely on the place of articulation information from the vowel preceding the fricative noise.

The crucial language-specific difference for fricatives was carried not by the formant transitions but by the first portion of the frication in VC syllables. On the basis of this, one might argue that listeners of languages with perceptually similar fricatives have refined their perception of cues in the fricative spectrum rather than relying on coarticulatory cues. What speaks against this is that Polish and Spanish listeners were not generally better in identifying fricatives: their performance in the last gate, after hearing the entire frication, was lower than the performance of Dutch and Italian listeners. However, Polish and Spanish listeners gained more information about the place of articulation from a shorter portion of the frication. This earlier uptake of information appears to be a compensation for the lack of spectral information to individuate perceptually confusable native fricatives in Spanish and Polish. Also, the phoneme monitoring study by Wagner et al. (2006) showed that mismatching vowel information in cross-spliced VCV sequences impedes identification for listeners with several similar fricative contrasts. Combined, these results imply that, compared to Dutch and Italian, Spanish and Polish listeners rely more on the coherence of the vowel portion with the fricative noise. It is possible that the construct of coherence takes into account a procedure similar to the measure suggested by Hedrick and Ohde (1993), an evaluation of the amplitude of the vowel relative to the frication, which may play a more important role for Polish and Spanish listeners than for Dutch and Italian listeners.

Another outcome of the present study is that language-specific differences were found in VC syllables and not in CV syllables. In CV syllables, all listeners gained the same amount of information from the vowel portion following the constriction. If listeners benefit more from the information which comes in last, as suggested by Mann and Soli (1991), then coarticulatory information in CV syllables may be generally more accessible. Listeners who need more information may benefit from cues announcing a phoneme, thus from information which can be disregarded by other listeners. Note that among the languages of the world, regressive place assimilation is more common than progressive place assimilation (Jun, 2004). This might reflect a perceptual bias to "look ahead" in the acoustic signal in order to get information for less robust phonological features, such as place of articulation, the cues to which have been shown to be less robust in noisy conditions (Miller and Nicely, 1955), and to be perceived worse than cues to manner in normal listening conditions (Warner et al., 2005).

In addition to this, for listeners with only disparate native fricatives, the crucial and effectual information seems to be contained in the onset portion of the frication, whether it is the first third (Smits et al., 2003; Warner et al., 2005) or a portion longer than 40 ms (Jongman, 1989) of the frication. Listeners who are in need of more information to specify place of articulation may rely on the coarticulatory information that is temporally closest to this essential information. Temporal separation of these two sources of information may affect the way this information is integrated on a pre-categorical level of processing. Early studies by Repp (1982) and Mann and Repp (1980) report such effects of temporal segregation: the insertion of a silent interval between fricative and vowel reduces the perceptual effect of the vowel.

The observed difference in informativeness of parts of the acoustic signal for fricative identification does not necessarily mean that listeners with perceptually robust fricatives are not sensitive to this coarticulatory information. This rather suggests that they are relatively inattentive to this systematic acoustic variation when processing speech. There are numerous examples of language-specific weighting of acoustic information (e.g., Crowther and Mann, 1992). For instance, there are cross-language differences in the sensitivity to transitional cues for the /r/-/l/ distinction between Japanese, German and English listeners (Iverson et al., 2003).

With regard to fricatives, Li et al. (2011) showed that English and Japanese listeners differ as to which acoustic parameters they rely upon when classifying the fricatives /s/ and /ʃ/ produced by children. An additional acoustic analysis of Japanese and English fricatives, produced by adults (Li et al., 2009), showed differences in the realization of these fricatives, also in the realization of /s/, the fricative that is common to both languages. Furthermore, Li and colleagues (Li et al., 2011) show and argue that these cross-language differences in realization lead to language-specific patterns of errors made by 2- and 3-yr old Japanese and English learning children. Two logical consequences of these studies are particularly interesting: the link between language-specific production and perception, as well as their impact in speech development. These two points, and their relevance for the present study, will now be briefly discussed.

Wagner et al. (2006) hypothesized that the presence of perceptually confusable fricatives in the native fricative inventory leads to language-specific reliance on acoustic information. The number of contrasts within an inventory will, together with other factors such as phonotactic constraints, lead to language-specific realizations of phonemes. The phoneme inventory thus contributes to the definition of language-specific articulatory "output constraints" (Manuel, 1987): speakers will aim for phoneme realizations that maintain all native categories as distinct as possible. Native listeners are exposed to these language-specific realizations, which in turn form the basis for language-specific patterns of reliance on acoustic information.

What happens during speech development can be seen as optimization of perception: infants learn which patterns of acoustic variation are relevant and necessary in the language that surrounds them (e.g., Maye et al., 2002). For fricative identification, it has been shown that children weight transitional information more than adult listeners do (Nittrouer, 1992; Nittrouer and Studdert-Kennedy, 1987), and it is only around the age of 6, that English infants show an adult-like pattern of reliance on frication noise and transitions (Nittrouer, 2002). They rely on the information in the noise spectrum for robust fricatives like /s/ and /ʃ/, and on both, noise spectrum and transitions, for less robust fricatives like /f/ or /θ/. If infants growing up with English as their native language selectively reduce their reliance on transitional information, then it follows that infants growing up with languages with only disparate fricative noises are likely to reduce their reliance on coarticulatory cues for fricatives altogether.

Language-specific realizations of phonemes play an important role in speech development. They lead to language-specific patterns of production errors, and form the basis for language-specific sensitivities to acoustic information. Once

these sensitivities are established, however, they are difficult to readjust, as is evidenced by the ample body of studies on L2-learning and -training (e.g., Bradlow *et al.* 1997; Iverson *et al.*, 2005). This may also explain why the present study did not find a difference between materials produced by a Dutch speaker (main experiment), and the materials produced by a Spanish speaker (control experiment). Adult listeners, when performing the experimental task, were either not sensitive to the language-specific realizations, or the Dutch and Spanish materials were not dissimilar enough to cause significantly different perceptual consequences.

It might seem somewhat surprising that more cross-language differences were not found in the present study. After all, the four languages vary in more aspects than just in their phoneme inventories: they have different syllable structures, language-specific phonotactic constrains and assimilation patterns. Although all languages tested have similar plosive inventories, their language-specific phonotactics will have equipped listeners with very different expectations about where a plosive could occur and which phonemes could surround it (e.g., Weber, 2002). The fact that no differences were found in the way listeners processed place information for plosives, suggests that phonotactic effects were limited in the used experimental paradigm, and implies that listeners processed the stimuli at a level before such phonological effects could take place.

In the experimental task participants were familiarized with the fact that the stimuli were not real words, and did not resemble words in either of the languages tested. It is possible that listeners treated these stimuli as non-speech. A task involving more lexical activation, be it with real words or quasi-words as materials, may attest more cross-language differences that can be attributed to language-specific phonotactics. Phoneme monitoring, and gating as well, can be seen as somewhat metalinguistic in nature, and thus may obscure some phonological or lexical factors. A previous study compared phoneme monitoring across listeners from five different languages, and has found no effect of language-specific phoneme frequencies (Wagner and Ernestus, 2008). The fact that these experimental paradigms, nevertheless, find cross-language differences suggests that language-specific processing starts operating at very early stages of sound processing, somewhere between auditory processing and phonological categorization.

This study has documented differences and similarities in the uptake of acoustic information for phoneme identification among four languages. Despite many differences between the linguistic systems of the languages tested similarities were found in the processing of vowel-plosive and plosive-vowel syllables. These similarities might reflect some generally preferred perceptual patterns. Nonetheless, this study also attested language-specific patterns in the uptake of information to place of articulation for fricatives, suggesting that language-specific perceptual optimization is partly a function of listeners' native phoneme inventories.

ACKNOWLEDGMENTS

Baayen, R. H. (**2008**). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge), Chap. 7, pp. 263–299.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

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

Benki, J. R. (**2001**). "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phonetics **29**, 1–22.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am **101**(4), 2299–2310.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: evidence from 18 languages," J. Phonetics **27**, 207–229.

Crowther, C. S., and Mann, V. A. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**(4), 769–773.

Dorman, M. F., Raphael, L. J., Liberman, A. M., and Repp, B. H. (**1975**). "Some maskinglike phenomena in speech perception," Haskins Laboratories Status Report on Speech Research, SR-42/43, pp. 265–276.

Dorman, M. F. Studdert-Kennedy, M., and Raphael, L. J. (**1977**). "Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues," Percept. Psychophys. **22**, 109–122.

Fujimura, O., Macchi, M. J., and Streeter, L. A. (**1978**). "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study," Lang. Speech **21**(4), 337–346.

Grosjean, F. (**1980**). "Spoken word recognition processes and the gating paradigm," Percept. Psychophys. **28**, 267–283.

Grosjean, F. (**1996**). "Gating," Lang. Cogn. Process. **11**, 597–604.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang. Speech **1**, 1–7.

Hedrick, M. S., and Ohde, R. N. (**1993**). "Effect of the relative amplitude of frication on the perception of place of articulation," J. Acoust. Soc. Am. **94**, 2005–2026.

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of fricative consonants," J. Acoust. Soc. Am. **33**, 589–593.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English vertical bar r vertical bar-vertical bar l vertical bar to Japanese adults," J. Acoust. Soc. Am **118**, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada R., Diesch E., Tohkura Y., Ketterman A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Jassem, W. (**1968**). "Acoustic description of voiceless fricatives in terms of spectral parameters," in *Speech Analysis and Synthesis*, edited by W. Jassem (Panstwowe Wydawnictwo Naukowe, Warsaw), pp. 189–206.

Jongman, A. (**1989**). "Duration of frication noise required for identification of English fricatives," J. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A. (**1998**). "Are locus equations sufficient or necessary for obstruent perception?," Behav. Brain Sci. **21**, 271–272.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Jun, J. (**2004**). "Place assimilation," in *Phonetically Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge), pp. 3.1–3.48.

Klaassen-Don, L. E. O. (**1983**). "The influence of vowels on the perception of consonants," Doctoral Dissertation, Leiden University.

Li, F., Edwards, J., and Beckman, M. E. (**2009**). "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," J. Phonetics **37**, 111–124.

Li, F., Munson, B., Edwards, J., Yoneyama, K., and Hall, K. (**2011**). "Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development," J. Acoust. Soc. Am. **129**, 999–1011.

Liberman, A. M., Delattre, P., Cooper, F. S., and Gerstmann, L. J. (**1954**). "The role of consonant-vowel transition in the perception of the stop and nasal consonants," Psychol. Monographs **279**(68), 1.

Lipski, S. C. (**2006**). "Neural correlates of fricatives across language boundaries," Doctoral Dissertation, University of Stuttgart.

Lisker, L. (**2001**). "Hearing the Polish sibilants [s, š, ś]: Phonetic and auditory judgments," in *To Honour Eli Fischer—Jørgensen Travaux du Cercle Linguistique de Copenhague, 31*, edited by N. Grommum and J. Rischel (Reitzel, Copenhagen), pp. 226–238.

Mann, V., and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [sh]-[s] distinction," Percept. Psychophys. **28**, 213–228.

Mann, V., and Soli, S. D. (**1991**). "Perceptual order and the effect of vocalic context on fricative perception," Percept. Psychophys. **49**, 399–411.

Manuel, S. (**1987**). "Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages," Ph.D. thesis, Yale University, New Haven, CT.

Martinez-Celdran, E., Fernandez-Planas, A. M., and Carrera-Sabate, J. (**2003**). "Illustrations of the IPA: Castilian Spanish," J. IPA **33**(2), 255–259.

Maye, J., Werker, J. F., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**(3), B101–B111.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonet. **20**, 351–382.

Nittrouer, S., Miller, M., Crowther, C., and Manhart, M. (**2000**). "The effect of segmental order on fricative labeling by children and adults," Percept. Psychophys. **62**(2), 266–284.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.

Pols, L. C. W., and Schouten, M. E. H. (**1978**). "Identification of deleted consonants," J. Acoust. Soc. Am. **64**, 1333–1337.

Recasens, D. (**1987**). "An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences," J. Phonetics **15**, 299–312.

Repp, B. H. (**1978**). "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants," Percept. Psychophys. **24**, 471–485.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psych. Bull. **92**, 81–110.

Reubold, U., Harrington, J., and Kleber, F. (**2010**). "Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers," Speech Commun. **52**, 638–651.

Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. l(27), 379–423, 623–656.

Smits, R. (**2000**). "Temporal distribution of information for human consonant recognition in VCV utterances," J. Phonet. **28**, 111–135.

Smits, R., ten Bosch, L., and Collier, R. (**1996**). "Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment," J. Acoust. Soc. Am. **100**, 3852–3864.

Smits, R., Warner, N. L., McQueen, J. M., and Cutler, A. (**2003**). "Unfolding of phonetic information over time: A database of Dutch diphone perception," J. Acoust. Soc. Am. **113**(1), 563–574.

Sorianello, P. (**2002**). "I suoni fricativi dell'italiano fiorentino" ("The fricative sounds of the Florentine Italian)," Quad. Lab. Linguist. **3**, 26–39.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1158–1872.

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (**1998**). "Linear correlates in the speech signal: The orderly output constraint," Behav. Brain Sci. **21**, 241–299.

Tartter, V. C., Kat, D., Samuel, A. G., and Repp, B. H. (**1983**). "Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions," J. Acoust. Soc. Am. **74**, 715–725.

Venables, W. N., and Ripley, B. D. (**2002**). *Modern Applied Statistics with S*, 4th ed. (Springer, New York).

Wagner, A. (**2008**). "Phoneme inventories and patterns of speech sound perception," Doctoral dissertation, Radboud University Nijmegen (Ponsen & Looijen bv, Wageningen), electronic version available at: http://repository.ubn.ru.nl/handle/2066/72855.

Wagner, A., and Ernestus, M. (**2008**). "Identification of phonemes: Differences between phoneme classes and the effect of class size," Phonetica **65**, 106–127.

Wagner, A., Ernestus, M., and Cutler, A., (**2006**). "Formant transitions in fricative identification: The role of native fricative inventory," J. Acoust. Soc. Am. **120**, 2267–2277.

Warner, N. L., Smits, R., McQueen, J. M., and Cutler, A. (**2005**). "Phonological and statistical effects on timing of speech perception: Insight from a database of Dutch diphone perception," Speech Commun. **46**, 53–72.

Warren, P., and Marslen-Wilson, W. D. (**1987**). "Continuous uptake of acoustic cues in spoken word recognition," Percept. Psychophys. **41**, 262–275.

Weber, A. (**2002**). "Assimilation violation and spoken-language processing: A supplementary report," Lang. Speech **45**, 37–46.

Whalen, D. H. (**1981**). "Effects of vocalic formant transitions and vowel quality on the English [s]-[s] boundary," J. Acoust. Soc. Am. **69**, 275–282.