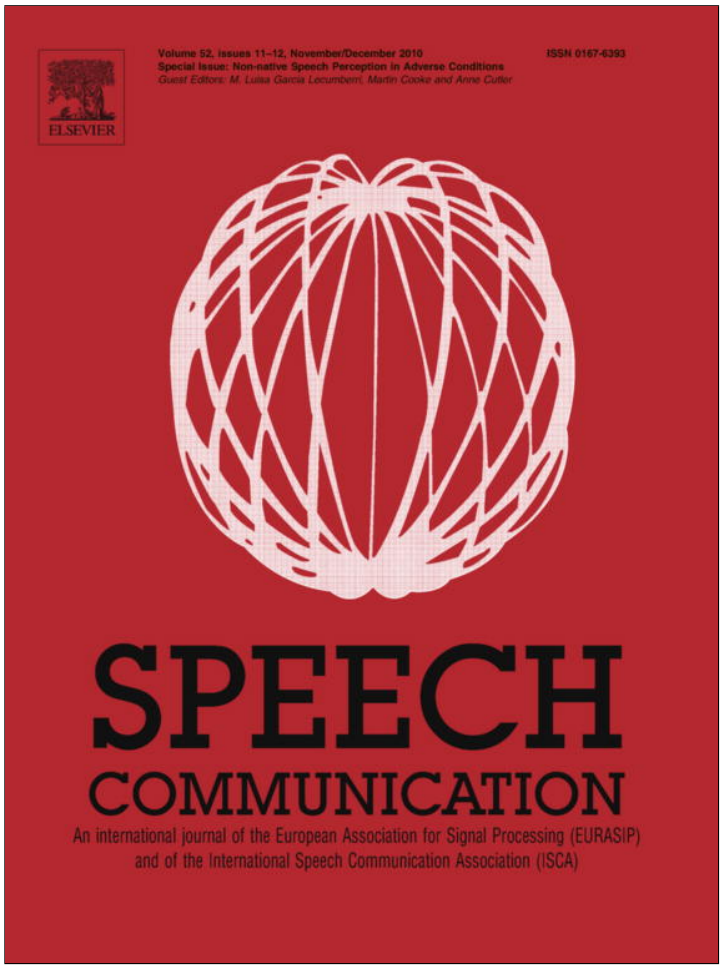

Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 954–967
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages

Martin Cooke [a,b,*], Maria Luisa Garcia Lecumberri [b], Odette Scharenborg [c], Wim A. van Dommelen [d]

[a] *Ikerbasque, Basque Foundation for Science, 48011, Bilbao, Spain*
[b] *Language and Speech Laboratory, Faculty of Letters, University of the Basque Country, Paseo de la Universidad, 5, 01006, Vitoria, Spain*
[c] *Centre for Language and Speech Technology, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands*
[d] *Department of Language and Communication Studies, NTNU, NO-7491, Trondheim, Norway*

Received 2 September 2009; received in revised form 12 April 2010; accepted 12 April 2010

Abstract

Processing speech in a non-native language requires listeners to cope with influences from their first language and to overcome the effects of limited exposure and experience. These factors may be particularly important when listening in adverse conditions. However, native listeners also suffer in noise, and the intelligibility of speech in noise clearly depends on factors which are independent of a listener's first language. The current study explored the issue of language-independence by comparing the responses of eight listener groups differing in native language when confronted with the task of identifying English intervocalic consonants in three masker backgrounds, viz. stationary speech-shaped noise, temporally-modulated speech-shaped noise and competing English speech. The study analysed the effects of (i) noise type, (ii) speaker, (iii) vowel context, (iv) consonant, (v) phonetic feature classes, (vi) stress position, (vii) gender and (viii) stimulus onset relative to noise onset. A significant degree of similarity in the response to many of these factors was evident across all eight language groups, suggesting that acoustic and auditory considerations play a large rôle in determining intelligibility. Language-specific influences were observed in the rankings of individual consonants and in the masking effect of competing speech relative to speech-modulated noise.

Keywords: Consonant identification; Non-native; Cross-language; Noise

1. Introduction

Understanding speech produced in a non-native language can be a challenging experience, especially in the less-than-ideal conditions typical of many communicative situations. Non-native speech perception is heavily influenced by a listener's first language (L1) sound system (Best, 1995; Flege, 1995; Kuhl, 1993). Between-language phonetic distance (Hazan and Simpson, 2000), language competence (Imai et al., 2005) and orthographic interference (Detey and Nespoulous, 2008) are other factors which contribute to intelligibility for non-native listeners (NNLs). However, the ease with which a spoken message is understood under adverse conditions is influenced by other factors too. Speakers differ in intrinsic intelligibility (Bradlow et al., 1996; Hazan and Simpson, 2000; Barker and Cooke, 2007) while different speaking styles make speech more or less likely to be understood (Bradlow and Bent, 2002; Picheny et al., 1985; Bradlow and Alexander, 2007). The type and level of the background noise is also relevant: competing speakers produce less masking in the auditory

* Corresponding author at: Ikerbasque, Basque Foundation for Science, 48011, Bilbao, Spain.
E-mail addresses: m.cooke@ikerbasque.org, M.Cooke@dcs.shef.ac.uk (M. Cooke).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.04.004

periphery than stationary or slowly-varying noise when matched for level (Festen and Plomp, 1990) but concurrent speech can also have a negative impact due to the informational masking effect of linguistic interference (Carhart et al., 1969; Brungart et al., 2001). Differences related to individual speakers, speaking styles and noise types affect not only non-native listeners but native listeners (NLs) too (Bradlow and Pisoni, 1999). However, what is currently unclear is whether such factors influence NLs and NNLs to the same extent. Determination of the relative contributions of language-independent and language-dependent processing is fundamental in understanding speech perception and has important consequences for applications such as speech enhancement, for instance, in predicting the likely benefit of specific signal treatments.

One approach to teasing apart the two types of processing is to compare NL and NNL responses on tasks which simulate adverse conditions such as added noise (Florentine et al., 1984; van Wijngaarden et al., 2002; Cutler et al., 2004), simulated reverberation (Takata and Nabelek, 1990; Rogers et al., 2006) or synthetic speech (Alamsaoutra et al., 2006). For practical reasons, most studies of this type have been limited to pairs of languages, usually with English as the target language, paired variously with French (Florentine et al., 1984), Japanese (Takata and Nabelek, 1990), Italian (Mackay et al., 2001), Dutch (Cutler et al., 2004) or Spanish (Mayo et al., 1997; García Lecumberri and Cooke, 2006; Rogers et al., 2006), among others. However, it is difficult to draw strong conclusions about language-independent factors affecting speech intelligibility from paired-language studies, since the reduction in performance of non-native listeners could be ascribed to either L1 influences or to acoustic factors. Examining differences with respect to native controls is insufficient since the NL group itself has its own L1 influences which affect intelligibility in adverse conditions. An example for English would be the orthographic influence on /θ,ð/ distinctions which may further complicate the acoustic-based confusion, possibly even more so than for some non-native listeners lacking this orthographic influence in their own language.[1]

Comparing NLs and a single NNL group may be valid when the issue is one of examining the effects of linguistic competence. However, two groups may be insufficient to assess language-independent factors, particularly if one of them has the additional advantage of native competence and the consequent extensive exposure to the stimuli in question. The use of several NNL groups processing the same target language may provide a clearer picture of language-dependent and independent influences.

Occasionally, three L1s have been compared (e.g. English–Spanish–Japanese in Hazan and Simpson, 2000; English–Spanish–Dutch in Cutler et al., 2008). In the latter study, Cutler et al. presented Dutch listeners with a subset of stimulus conditions previously tested with English and Spanish listeners by García Lecumberri and Cooke (2006) to address the issue of whether a disproportionate effect of noise relative to quiet on Spanish listeners found by García Lecumberri and Cooke (2006), but not for Dutch listeners in (Cutler et al., 2004), was due to listener or stimuli differences. The Cutler et al. (2008) study demonstrates the potential value of using a wider range of listener groups differing in L1 in understanding the specific contributions of task and other factors to both native and non-native speech perception.

The current study extends this multi-language approach to eight European languages. All listeners undertook the same task, which was to identify consonants in vowel–consonant–vowel (VCV) tokens produced by British English speakers in a range of additive noise conditions. While in principle the data gathered from multi-language studies can be used to address issues involving detailed explanations for response patterns for specific L1 groups, the main focus of the current study was to investigate and measure responses across groups of listeners with different L1s in order to distinguish language-independent from language-specific processing in speech perception.

The key issue for the current study is the extent to which different L1 groups respond similarly to variables such as speaker (Bradlow and Pisoni, 1999) and vowel context (Dubno and Levitt, 1981; Hazan and Simpson, 2000; Jiang et al., 2006). For example, are consonants in some vowel contexts more intelligible for all listener groups regardless of L1? If some vowel contexts are favoured by all groups, then an acoustic basis for the preference is more likely than an explanation based on a learned relationship which ought to be sensitive to the L1 of a listener group. Noise is known to affect the rôle played by specific perceptual cues (Wang and Bilger, 1973; Hazan and Simpson, 2000; Parikh and Loizou, 2005; Jiang et al., 2006; van Engen and Bradlow, 2007) so it might be expected that, for instance, the ranking of speaker intelligibility will be masker-dependent. Here, we examine the similarity of listener group responses for to the following factors: (i) noise type, (ii) speaker, (iii) vowel context, (iv) consonant, (v) phonetic feature classes, (vi) stress position, (vii) gender and (viii) stimulus onset relative to noise onset.

2. Methods

2.1. Listener groups

Listeners were speakers of one of eight European languages. As well as a native group of British English listeners (en), native listeners of Czech (cz), Dutch (du), German (ge), Italian (it), Norwegian (no), Romanian (ro) and Spanish (sp) participated. This set of languages, four Germanic, three Romance and one Slavic, was chosen on the pragmatic basis that countries where these languages are spoken hosted participants in the Marie Curie Research

[1] Other than English, of the languages tested in the current study only Spanish has these sounds, and then very rarely in positions where an orthography-based influence might arise. For Spanish listeners, English /ð/ is typically confused with /d/.

Training Network "Sound to Sense", with the exception of the German group. Listeners were students and staff at the following institutions: University of Sheffield, Charles University Prague, Radboud University Nijmegen, University of Oldenburg, University Federico II (Naples), Norwegian University of Science and Technology (Trondheim), University of Cluj-Napoca and the University of the Basque Country (Vitoria). All listeners were tested in their own institutions. They were asked to report if they were aware of having any hearing problems and to rate their level of English on a 4-point scale. Table 1 summarises listener group information. A Kruskal–Wallis rank sum test indicated that the distribution of self-assessed competence was not the same for each group ($c^2(6,\ N = 465) = 54.0$, $p < 0.001$). We examine the relationship between self-assessed competence and performance in this task in Section 3.2.

2.2. Speech material

Speech material was drawn from an existing corpus of British English intervocalic consonants (VCVs) spoken in a number of vowel and stress combinations (Cooke and Scharenborg, 2008), recorded at the University of Sheffield. This corpus contains VCV tokens formed from all 24 consonants of British English (/p, b, t, d, k, g, tʃ, dʒ, f, v, θ, ð, s, z, ʃ, ʒ, h, m, n, ŋ, l, r, j, w/) in the context of all nine combinations of the vowels /i:, u:, æ/ for both front and end stress (e.g. /ˈæbæ/ versus /æˈbæ/). Note that while these patterns are not all valid combinations for English (the long/tense vowels cannot be followed by the velar nasal), speakers were aware that they were producing nonsense tokens. The current study used VCV tokens from four male and four female talkers.

2.3. Maskers

Listeners identified VCVs in quiet and six different additive noise backgrounds but the current analysis focuses on a representative subset of noise conditions, viz. speech-shaped noise (SSN), speech-modulated noise (SMN) and competing speaker (CS). The three noise backgrounds were selected on the basis that they provide different types of spectral and temporal masking and because they permit the rôles of informational and energetic masking to be examined. Energetic masking refers to the effect of interactions between target and masker in the auditory periphery, while informational masking is a cover term for the effect of more central processes which reduce intelligibility still further once energetic masking has been taken into account. Speech-shaped noise is a pure energetic masker with a fixed spectrum and no significant temporal modulations. Speech-modulated noise shares its spectral shape with SSN but has temporal envelope modulations derived from natural speech. SMN is also a pure energetic masker, containing no intelligible components, but it differs from SSN in permitting occasional clear glimpses across the entire spectrum of any signal mixed with it. A competing speaker contains significant modulations in both frequency and time and produces both energetic and informational masking since audible components of the masker can compete with those of the target, both for a listener's attention and in the assembly of speech fragments into a coherent stream.

Masker signals for each of the three noise conditions were derived from speech material from eight talkers (four male, four female) drawn from an existing corpus (Lu, 2010). Speech-shaped noise was generated by passing white noise through a 50 coefficient filter derived from the LPC spectrum resulting from the sum of 200 sentences. Competing speech maskers were randomly-chosen segments from the sentence material. As a consequence, the four male and four female background talkers were equally-likely to be chosen as maskers. Two of the female talkers who contributed the background material were among the set whose VCV tokens were used, so on about 3% of trials the foreground and background talkers were identical. To generate speech-modulated noise maskers, envelopes from random segments of competing speech were multiplied sample-wise with fragments of SSN.

Table 1
Listener group data. These figures are based on the 178 listeners whose responses were analysed in the paper and does not include participants who were removed from the analysis as outliers after taking part in the tests. SAC indicates self-assessed competence level in English based on the scale 1 = basic, 2 = intermediate, 3 = advanced, 4 = fluent.

Native language	*N*	Age: mean (SD)	SAC: mean (SD)
English	23	29 (8.3)	
Czech	18	21 (2.9)	2.44 (0.60)
Dutch	15	30 (9.0)	3.07 (0.69)
German	18	26 (3.5)	2.33 (0.67)
Italian	18	28 (5.3)	2.44 (0.60)
Norwegian	21	23 (2.8)	2.95 (0.66)
Romanian	26	23 (0.6)	2.58 (0.80)
Spanish	39	22 (4.9)	2.77 (0.48)

2.4. Stimuli

Each test condition involved the identification of a set of 384 VCV tokens, made up of one front-stressed and one end-stressed instance of each of the 24 consonants from each of eight speakers. Vowel contexts were chosen randomly and consequently were approximately uniformly-distributed across the test material. In the masked conditions, tokens were added to noise signals of 1200 ms duration. The onset time of the VCV token relative to the noise was varied in order to make the appearance of the target item unpredictable within the noise, since it has been suggested that predictability may benefit native listeners more than non-natives (Cutler et al., 2008). Onsets took one of eight values linearly-spaced in the range 0–400 ms. Onset times were balanced so that each consonant occurred the same number of times at each of the eight onsets. For each

token, the noise signal was scaled to produce a global SNR of −6 dB in the region where the token was present. This SNR value was chosen on the basis of pilot tests and previous experiments to produce consonant identification rates avoiding floor and ceiling effects for both native and non-native listeners (García Lecumberri and Cooke, 2006).

2.5. Perception tests

Listener groups were tested in eight countries using the same presentation software and following similar testing instructions. Tests were carried out in quiet laboratories or booths. Listeners identified consonants by selecting from a grid on a computer screen. Each consonant was represented by a common English grapheme combination and sample word (see Fig. 1). The use of graphemes was considered necessary since the majority of listeners were unfamiliar with phonetic symbols. This choice represents a compromise between testing a normal population via spelling, increasing the chances of orthographic influences and ambiguities, or testing phonetically-trained but unrepresentative listeners with phonetic symbols. In either case, the task is not a pure perceptual one but also metalinguistic to some degree. Listeners were given an explanation of the sound-grapheme correspondences and if a particular example word was felt to be confusing for a specific language group (due to cognates, for instance), a different example word was chosen.

Listeners underwent a short practice session containing 72 stimuli in quiet (three examples of each consonant) prior to the main test. During practice, feedback was provided for incorrect answers, and participants were able to listen to these stimuli as many times as required. For the main test, the quiet condition was presented first, followed by the noise conditions in randomised order. Stimuli were presented at a comfortable listening level. Since the quiet condition was always the first to be presented, it served as an extensive familiarisation test for listeners. In fact, response times were longer in quiet than in subsequent masked conditions for all listener groups, including natives, sometimes by substantial amounts. Consequently, here we focus solely on the noise conditions.

2.6. Selection of listeners and responses

In total, 207 listeners participated across the eight countries. Excluded from the sample were four listeners who reported hearing problems, three who were not natives of the language group and nine who did not complete all the conditions. Although most listeners were young adults, an analysis of age distributions for each language group revealed some variation across languages with one group having a long tail of older listeners. To reduce the imbalance, three listeners aged over 50 were removed from subsequent analysis. An initial analysis of consonant identification rates revealed that ten listeners were outliers in one or more conditions. These listeners were excluded, leaving a total of 178 listeners for further analysis. Also excluded were individual responses with response times greater than 10 s. Following data and listener selection, 270268 responses remained.

3. Results

3.1. Response time analysis

Response time (RT) is widely used in the field of human speech processing as a measure of relative processing difficulty (Cutler and Norris, 1979), including for the task of phoneme detection (Foss and Blank, 1980; Cutler et al., 1987). For the noise conditions in the current study, the VCV onset time within the noise was subtracted from the overall response time to produce a corrected measure. Measured across all listeners, weak but significant negative correlations between identification performance and corrected RT exist for each masker condition (CS: $r = -0.2$, $t(176) = -2.7$, $p < 0.01$; SMN: $r = -0.24$, $t(176) = -3.2$,

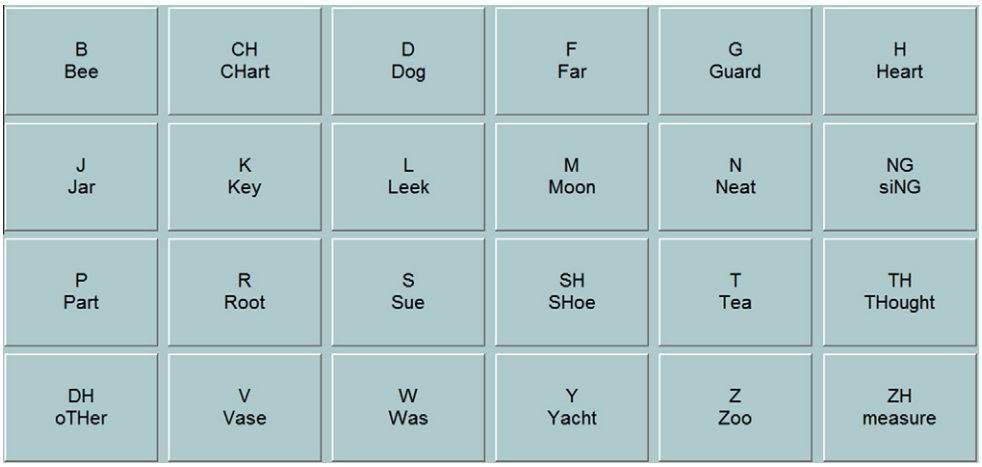

Fig. 1. Screenshot of perception testing software.

$p < 0.01$; SSN: $r = -0.15$, $t(176) = -1.97$, $p = 0.051$). Thus, for all test conditions, shorter response times correlate with a better identification performance, although the correlation is marginal in the case of the SSN masker.

Mean response times separated by correctness of response are displayed in Fig. 2 for each language group and noise condition. To explore differences among the noise backgrounds, separate mixed-effects analyses (Pinheiro and Bates, 2000; Pinheiro et al., 2008) were carried out for correct and incorrect responses, with L1 group and noise type as fixed effects and listener as a random effect. Neither correct nor incorrect responses led to a significant interaction between noise type and language (correct: $p = 0.18$; incorrect: $p = 0.28$). In both cases language was a significant main effect (correct: $F(7,170) = 2.79$, $p < 0.01$; incorrect: $F(7,170) = 4.86$, $p < 0.001$). However, this was due solely to the fast responses from the English and Dutch groups. While one might expect native listeners to respond more rapidly than non-native groups, it is not clear why Dutch listeners were so much faster than the other six non-native listener groups. Comparing all L1 groups except the English and Dutch, there was no significant language effect for correct ($p = 0.65$) or incorrect ($p = 0.24$) responses. For correct responses, there was a small effect of noise background ($F(2,268) = 3.68$, $p < 0.05$) although subsequent analyses showed no differences between pairs of noise types. No effect of noise type was seen in response times for incorrect responses ($p = 0.17$).

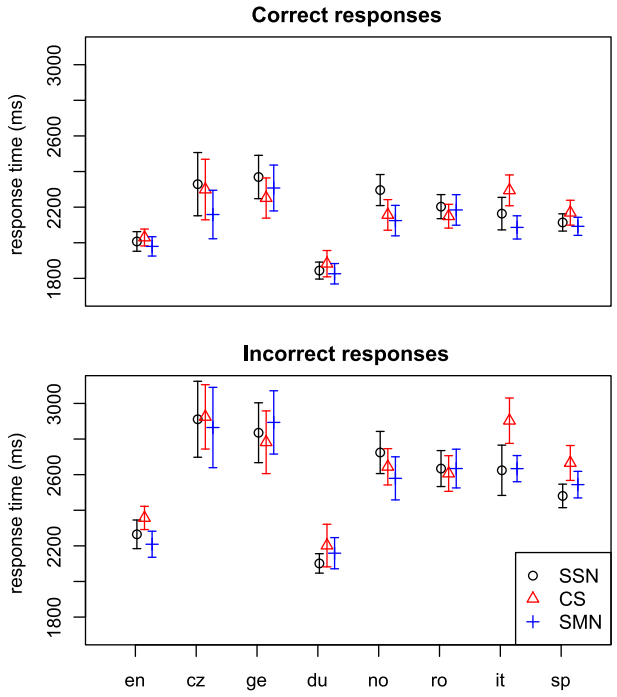

Fig. 2. Mean response times for each masker condition and language group for correct and incorrect responses. Here and elsewhere the ordering of listener groups is based on their mean consonant identification rate in the noise conditions, and vertical bars indicate ±1 standard error.

3.2. Consonant identification scores

Fig. 3 displays overall recognition performance as percentage correct consonant identification for each masker and language group. Of interest for the theme of the current study is the finding that the intelligibility ranking of test conditions is somewhat similar across the language groups tested here. Listener performance for SSN was always worse than the two modulated maskers (SMN and CS). While the CS masker produced similar or greater masking than the modulated noise, the degree of difference between the two showed some language-specific influences which we examine below.

Percentage correct consonant identification scores were converted to rationalised arcsin units (RAU; Studebaker (1985)) and subjected to a linear mixed-effects analysis with language group and test condition as fixed effects and listener as a random effect. This confirmed highly significant main effects of test condition ($F(2,340) = 324$, $p < 0.001$) and listener L1 ($F(7,170) = 32.1$, $p < 0.001$), as well as a significant interaction between the two ($F(14,340) = 2.10$, $p < 0.05$). The interaction is entirely due to differential responses to the SMN and CS maskers.

When listeners are aggregated into L1 groups, and excluding native listeners, there is no significant correlation between mean self-assessed competence and mean score in this task for any masker background (min $p = 0.56$). Even at the level of individual listeners we find no significant correlation between self-assessed competence and mean consonant identification scores across all noise conditions ($r = 0.06$, $p = 0.21$) or for any individual noise condition (max $r = 0.09$, min $p = 0.29$). Thus, self-assessed competence in level of English is not in any way a useful predictor of scores on this task.

3.3. Transmitted information analysis

Transmitted information analysis (Miller and Nicely, 1955) provides a more detailed picture of the influences

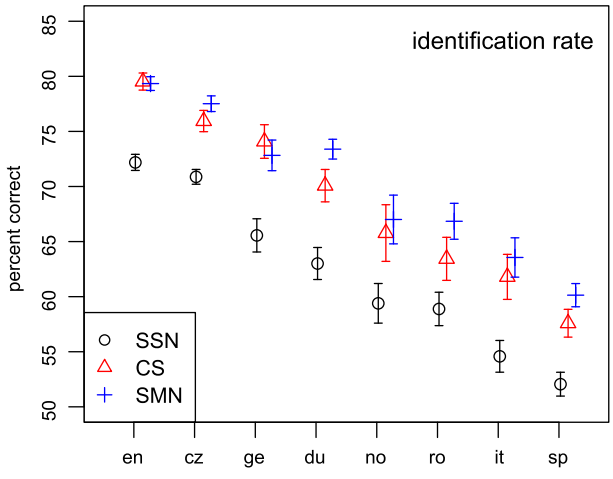

Fig. 3. Consonant identification rates for each masker and language group.

Table 2
Feature definition. Place: bi, bilabial; al, alveolar; ve, velar; fr, fricative; pa, postalveolar; la, labiodental; de, dental; gl, glottal; pl, palatal. Manner: pl, plosive; af, affricate; fr, fricative; na, nasal; li, liquid; ap, approximant. Sibilant: ir, irrelevant.

Consonant	p	b	t	d	k	g	tʃ	dʒ	f	v	θ	ð
Voice	−	+	−	+	−	+	−	+	−	+	−	+
Place	bi	bi	al	al	ve	ve	pa	pa	la	la	de	de
Manner	pl	pl	pl	pl	pl	pl	af	af	fr	fr	fr	fr
Sibilant	ir	ir	ir	ir	ir	ir	+	+	−	−	−	−
	s	z	ʃ	ʒ	h	m	n	ŋ	l	r	j	w
Voice	−	+	−	+	−	+	+	+	+	+	+	+
Place	al	al	pa	pa	gl	bi	al	ve	al	al	pl	bi
Manner	fr	fr	fr	fr	fr	na	na	na	li	li	ap	ap
Sibilant	+	+	+	+	−	ir	ir	ir	ir	ir	ir	ir

of specific phonetic features such as place or manner on consonant identification in noise, measured as the proportion of information for a given feature which is available to the listener (see Chapter 10 of Loizou, 2007 for a detailed example). Table 2 lists the features and their attributes used in the current study.[2]

Fig. 4 shows the proportion of information transmitted for the features voice, place, manner and sibilant. Voice is the least well-transmitted feature for all language groups and maskers. In addition to poor transmission of voicing, it is notable that SSN adversely affects the transmission of manner information, probably because a stationary masker lacks the temporal fluctuations important for distinguishing manner classes such as plosives and fricatives. Sibilance survives stationary noise better than the other features, most likely due to the greater presence of energy at high frequencies for the sibilants allowing an escape from the masking effects of speech-shaped noise at these frequencies. In fact, all noise backgrounds tested here show similar transmission of the sibilant feature, suggesting that the equivalence of their long-term spectrum gives rise to equally-likely glimpses of high frequency spectral information important for sibilant consonants.

A linear mixed-effects analysis with feature, test condition and language as fixed effects and listener as a random effect confirmed the above findings, in particular demonstrating a clear interaction effect of different noise backgrounds on the transmission of phonetic features ($F(6,1530) = 61.6$, $p < 0.001$). The three way interaction (language × feature × test set) was not significant ($p = 0.98$), but all two way interactions were significant at the $p < 0.001$ level. Apart from the feature × test set interaction mentioned above, the interaction of language and test set is partly due to the differences in response to the competing speaker across languages (see Section 3.4), while the language × feature interaction appears to stem from between-language variability in voicing transmission. Importantly, while the effect of features differed across test conditions, this variation was similar for all language groups, given the lack of a 3-way interaction. For example, sibilant becomes the best transmitted feature in SSN for all groups, and manner suffers more than place in SSN relative to other noise conditions, and again for all listener groups. These findings highlight language-independent effects of different masker types on phonetic feature transmission.

3.4. Effect of masker intelligibility

Differences in consonant identification scores in the competing speaker and speech-modulated noise conditions can be used to explore differential effects of intelligible versus unintelligible maskers, effects which constitute one aspect of informational masking. Fig. 3 suggests an overall tendency for listeners to suffer less in the presence of SMN than for the CS masker, a difference confirmed by a mixed-effects analysis ($F(1,170) = 24.0$, $p < 0.001$). However, not all listener groups behave in the same way with respect to these maskers, as evidenced by a significant language by masker interaction ($F(7,170) = 2.96$, $p < 0.01$). Of particular interest is the finding that the masking effect of SMN and CS was identical for the native listener group ($t(22) = 0.46$, $p = 0.65$) with a difference in consonant identification rates of less than 0.2%. Four of the seven non-native groups showed significant or near-significant adverse effects of the competing speech masker (du: $t(14) = -3.36$, $p < 0.01$; ro: $t(25) = -2.68$, $p < 0.05$; it: $t(17) = -1.83$, $p = 0.08$; sp: $t(38) = -3.6$, $p < 0.001$). At the level of individual listeners, there was a significant negative correlation between mean score in noise and the adverse impact of the CS masker relative to SMN (including the native group: $r = -0.31$, $t(176) = -4.37$, $p < 0.001$; excluding native listeners: $r = -0.27$, $t(153) = -3.48$, $p < 0.001$), suggesting that better performance on the task helps in combatting the competing speech masker.

The reduced non-native performance for CS compared to SMN in these cases is largely due to poorer reception of place and to a lesser extent sibilant features. The lack of impact on manner is not surprising since both maskers have temporal dips which help distinguish manner classes via relatively long-duration cues such as frication, nasality

[2] Transmitted information analysis was carried out using the FIX program available from ftp://pitch.phon.ucl.ac.uk/pub/fix/, stopping after the first iteration, which is equivalent to the Miller and Nicely (1955) procedure.

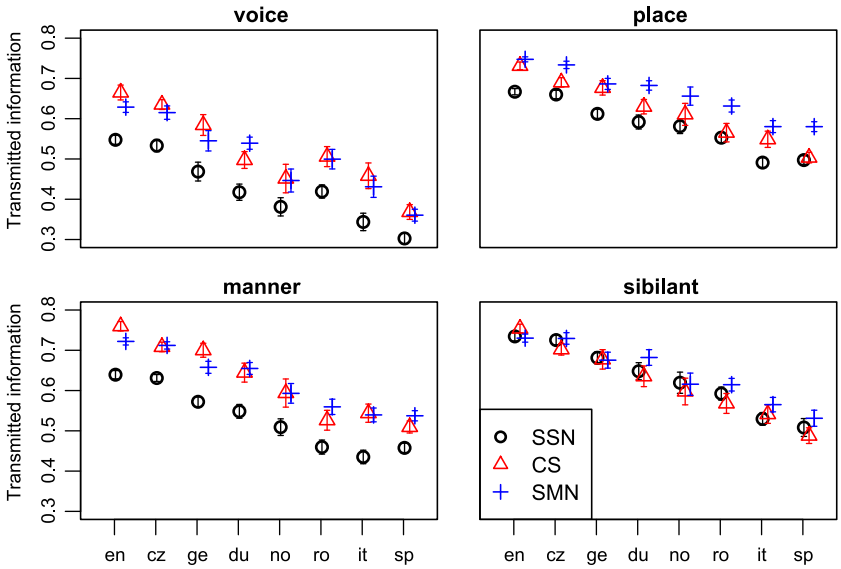

Fig. 4. Proportion of transmitted information for the features voice, place, manner and sibilant for each masker and language group.

and glide transitions. On the other hand, if the fine spectral detail which differentiates CS and SMN maskers were important, one might expect this to have an impact on the transmission of the voice feature. However, the relationship between the phonological category of voice and the acoustic signal is multi-faceted (Jiang et al., 2006), and it is likely that non-spectral cues to voicing, such as duration, are well-transmitted by both types of modulated noise.

3.5. Effects of stimulus variability

The previous sections investigated the effect of noise on consonant identification for native and non-native listeners, and the differential availability of features in different maskers. In the following analyses, we examine the degree of influence of contextual and paralinguistic factors (i.e. sources of stimulus variability). For example, does vowel context play a similar rôle in each language? Are some speakers more intelligible and others less intelligible for all listener groups? The degree of language-independent processing is assessed here in two ways. First, the effect size and ranking of levels within each factor is compared across language groups. Second, a quantitative measure of between-language agreement with respect to each factor is used to provide a numerical indication of the language-independence of each factor. We investigate the rôle of speaker, vowel context, consonant, gender, stress position and stimulus onset relative to noise onset. We also examine the similarity of response to the phonetic feature categories introduced in the previous section. These factors are not completely independent: speaker and gender are of course related, and phonetic feature transmission is based on consonant confusions. Nevertheless, examination of how levels within each factor are ranked by each listener group has the potential to provide insights into the scale of language-independence of these factors.

3.5.1. Univariate effect sizes

Similarity across the eight languages can be visualised using design plots (Figs. 5–7). Comparison of such plots across listener groups reveals similarity or otherwise both in the effect of different factors and in the ordering of levels within factors. Design plots also allow comparison of the influence of a factor with respect to other factors. For example, the variability in intelligibility across a range of speakers can be directly compared with that induced by changes in vowel context. Figs. 5–7 display mean scores for each of the levels of the factors consonant, vowel context, speaker, onset time, gender and stress position, for each L1 group. For brevity, these are averages across the three maskers but the effect of each masker is quantified separately in Section 3.5.2.

Taken together, the design plots in Figs. 5–7 suggest that the ranking of factor influence is virtually the same across languages. Of the six factors displayed, consonant identity has the largest effect on intelligibility for all listener groups. For example, individual consonant scores occupy a range from around 36% to 94% for native listeners. Vowel context is the next most influential factor for most groups, followed by speaker identity, stimulus onset and gender. Stress position has almost no effect for most groups, and only a small influence for Dutch and Spanish listeners, who identify consonants slightly better when presented in VCVs with end stress.

The larger influence of individual consonants on intelligibility is unsurprising since this is the area where L1 influences are most likely to be felt. For example, /z/ is poorly identified by the Spanish group but near or at the top for most of the other languages; /ʒ/ is the most intelligible consonant for Czech listeners but least intelligible for the Spanish group; German listeners have particular problems with /w/. These cases are clear examples of L1 influences, either from orthography or differences in phonemic inventories (e.g. lack of voiced fricative phonemes in Spanish).

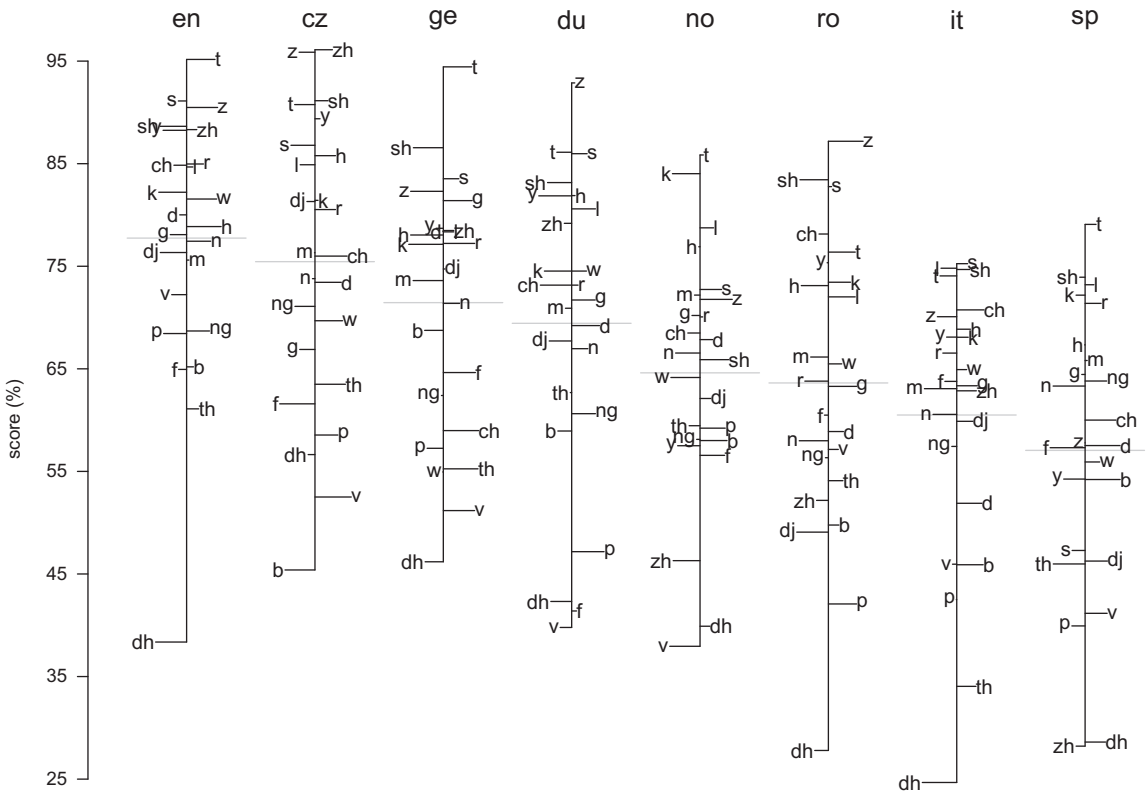

Fig. 5. Effect sizes for consonants averaged over noise conditions.

However, there are also indications of language-independent influences: the sibilant fricatives /s,z,ʃ/ and the plosive /t/, whose burst resembles the sibilants' profile, tend to be well-recognised by most language groups; weak fricatives /θ,ð,f,v/ and /p/ were particularly problematic for most listeners.

Speakers f1 and m1 are among the most intelligible, while speaker f3 is the most difficult for nearly all language groups. Contexts whose second vowel is /æ/ produce highest consonant scores for most groups. VCVs of type /i:_i:/ and /u:_u:/ led to the lowest scores. All groups scored more highly for the female speakers apart from the Spanish who score at equal rates for each gender. The difference, of about two percentage points, is remarkably consistent across non-native groups and is similar to the difference for native listeners. While the small number of speakers precludes firm conclusions, this finding echoes earlier studies which found an intelligibility benefit for female speakers in quiet (Bradlow et al., 1996; Hazan and Simpson, 2000; Hazan and Markham, 2004) and noise (Barker and Cooke, 2007).

The influence of stimulus onset relative to the noise reveals some interesting cross-language trends. The ranking of onset points is nonmonotonic for all listener groups but shows some clear commonalities. Onset points 1,3 and 7 are always ranked below average, while point 4 is top or near the top for all groups. Onset point 1, corresponding to a stimulus onset time of 57 ms into the noise, is particularly detrimental to most listener groups. The co-gated conditions (coded as 0) results in better than average consonant identification. The between-language similarities here are striking, suggesting at least some basis in low-level auditory processing. Forward masking by noise is maximally-disruptive immediately after onset (Moore, 2004), and it is notable that by the next onset point (104 ms, coded 2) performance is above average.

3.5.2. Between-language agreement

A quantitative measure of the extent to which a group of languages respond similarly to factor levels is provided by Kendall's coefficient of concordance, W (Kendall, 1948), which is often used as an index of inter-rater reliability (here, language groups can be treated as raters). Coefficients calculated using the function "kendall" from R's "irr" package (Gamer et al., 2007) are listed in Table 3 for the six factors considered above and for the additional factor of phonetic feature. Computation of Kendall's W used percentage correct scores for all factors apart from phonetic feature, where the proportion of transmitted information was used.

Nearly all of the coefficients of between-language similarity are highly significant, with a single exception: stress judgements for the SMN condition. Column means give an indication of how much between-language agreement exists for the various factors. Least similar are cross-language responses to consonants, which is to be expected

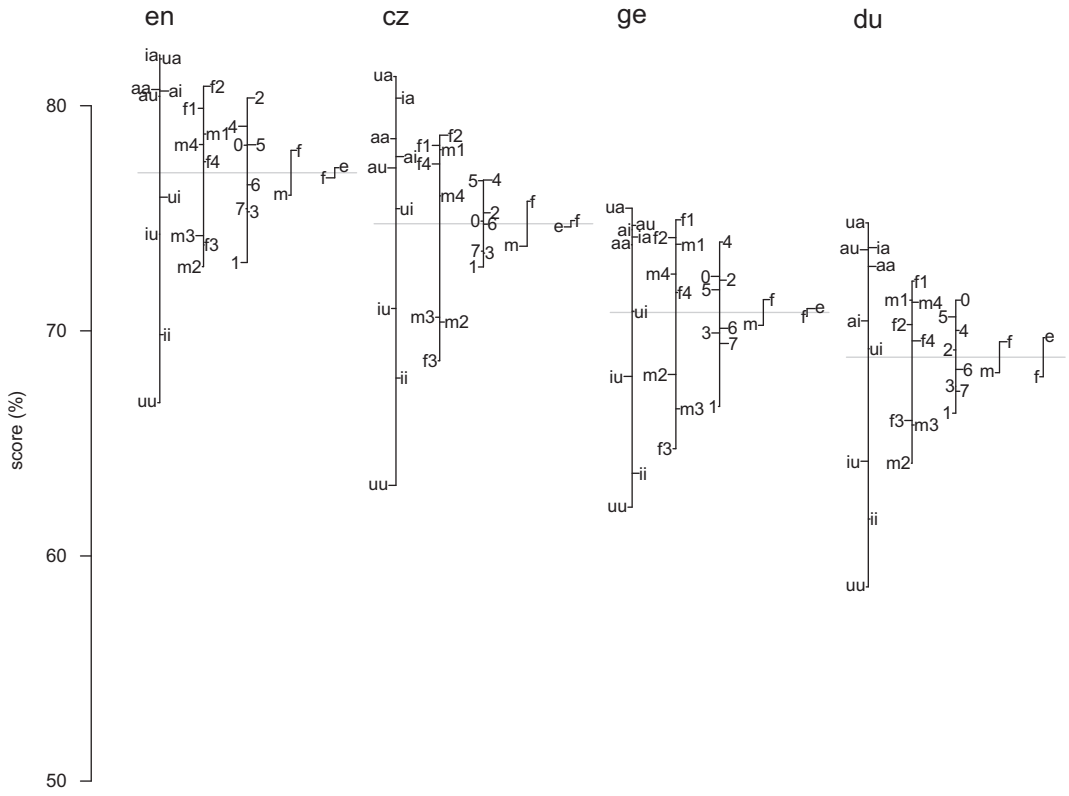

Fig. 6. Effect sizes for the factors (i) vowel context (e.g. ai indicates /æ_i:/); (ii) speaker (m1–4 are males, f1–4 are females); (iii) stimulus onset relative to noise onset (0 indicates co-gated, and each integer represents a further 57 ms gate, with 7 indicating 400 ms); (iv) gender and (v) stress position (f: front, e: end). Values are raw identification scores in percentages averaged over the three masker conditions. Mean performance in each group is indicated by a horizontal line across all factors. This plot is for English, Czech, German and Dutch.

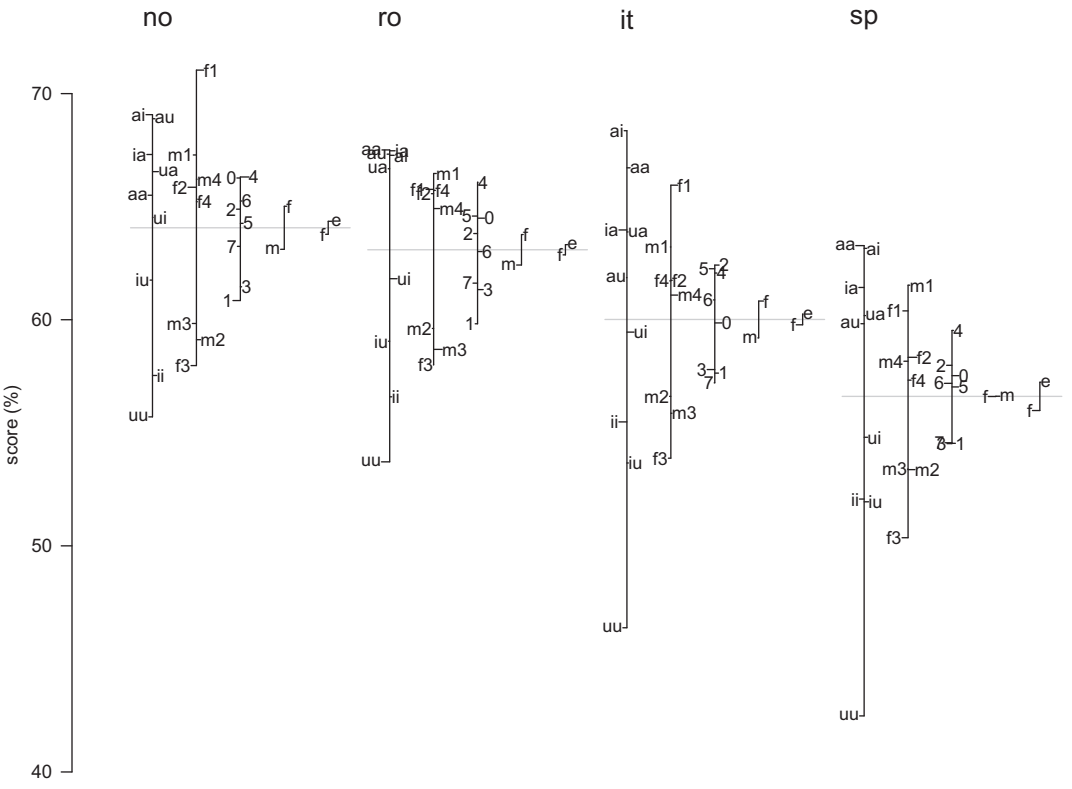

Fig. 7. As Fig. 6 but for Norwegian, Romanian, Italian and Spanish.

Table 3
Between-language agreement (Kendall's W coefficient of concordance) across the eight language groups for the factors listed in the columns.

	Speaker	Gender	Vowel context	Stress position	Consonant	Phonetic feature	Onset relative to noise	Mean (significant)
SSN	0.77***	1.00**	0.91***	1.00**	0.74***	1.00***	0.62***	0.84
CS	0.88***	0.56*	0.71***	1.00**	0.67***	0.64***	0.82***	0.72
SMN	0.91***	1.00**	0.70***	0.06 n.s.	0.67***	0.96**	0.86***	0.85
Mean	0.86	0.85	0.77	1.00	0.69	0.87	0.77	

n.s.: not significant. Row means are computed over all factors apart from stress.
* Significant at <0.05
** Significant at <0.01.
*** Significant at <0.001.

on the basis of L1 interference in the perception of similar sounds. However, at the level of phonetic factors a great deal of similarity is present, reflecting the phonetic feature ordering seen earlier in Fig. 3. The influence of vowel stress and gender was identical for all languages in two of the three masking conditions, bearing out the visual impressions in Figs. 6 and 7. However, some caution is needed to interpret findings based on these factors since they possess only two levels and are less robust. For instance, leaving out the native listener group leads to an increase in the gender coefficient for the CS masker from 0.56 to 1.0, but made very little difference to the other coefficients. Judgements of individual speaker intelligibility also showed high agreement across language groups. The influence of vowel context was somewhat less similar, as was onset time of the VCV token.

Mean coefficient values across all features apart from stress indicate that the two pure energetic maskers (SSN and SMN) led to similar between-language agreements (0.84 and 0.85) respectively, while the response to the competing speaker was less homogeneous (0.72). Leaving out the binary factors gender and stress reduces the difference between the maskers to some extent (SSN: 0.81, CS: 0.74, SMN: 0.82). Of the remaining five factors, the reduced agreement for CS relative to the other masker types is largely due to disagreements in the rankings of phonetic features.

4. Discussion

4.1. Degree of language-independence

A significant amount of similarity in factor rankings was observed throughout the study, highlighting a large degree of language-independence in speech perception. The ranking of transmitted information proportion for phonetic features was similar for most listener groups (e.g. all groups found voice to be the most poorly transmitted feature in noise). Similar changes in the rankings of phonetic features across test conditions were observed across listener groups. A great deal of similarity can also be seen in the influence of individual factor and levels within factors for speaker intelligibility, vowel context, token onset, gender and stress. These findings suggest that acoustic and auditory factors have a dominant influence in determining perceptual responses even across groups with different native languages. Unsurprisingly, the between-language similarity of responses to individual consonants, which reflects the degree to which different languages agree in the ease or difficulty of identifying consonants, is the factor most affected by L1 influences.

The fact that most listener groups agreed on the most and least intelligible speakers suggests that the factors which underlie speaker clarity are to some degree independent of L1, with perhaps only a minor rôle for aspects such as accent familiarity, where one would expect to see a difference between native and non-native rankings. However, in the current study, speakers all adopted a similar speaking style and speech rate, but it is worth noting that when deliberately "clear" speech is compared to conversational speech, the intelligibility gain in noise is substantially greater for native listeners than non-natives (Bradlow and Bent, 2002).

Regarding vowel context, our finding that /æ/ is a favourable context in noise, perhaps because of the presence of clear transitions cueing place and voicing, agrees with earlier studies (Hazan and Simpson, 1998; Hazan and Simpson, 2000; Jiang et al., 2006; Dubno and Levitt, 1981). Stress position had little effect but where there was an influence it favoured a slight overall benefit of end stress. In English, the target language in our study, front stress in bisyllabic words is a more natural pattern than end stress. Different realisations are associated with consonants appearing in stressed or unstressed syllables. In general, consonants have more distinctive or stronger realisations in stressed syllables (e.g. aspiration for voiceless plosives, complete closure for voiced plosives), which could account for stress preference in noise for some listener groups towards a pattern that enhances distinctiveness.

On the surface, our finding that co-gating the target and noise did not result in poorer performance than the cases with a positive noise lead time is at odds with previous findings on the effect of noise lead time on consonant identification where co-gating was shown to produce poorer identification scores than continuous noise (Ainsworth and Meyer, 1994) or noise with a lead time of 200 ms or longer (Cervera and Ainsworth, 2005). Those studies, however, did not test noise lead times as short as 50 ms and furthermore used plosive-vowel syllables rather than the more

extensive set of VCVs used here. The precise time-course of leading noise on consonant perception merits further study.

By quantifying the degree of between-language agreement, it is possible to compare the influence of several factors with respect to native language influence on consonant perception. Here, stress position, gender, speaker and phonetic feature (voice, place, manner, sibilant) showed relatively little L1 influence while vowel context and token onset time afforded somewhat more. Further, the effect of masker type on L1 influence can be compared. Here, SSN and SMN, both regarded as pure energetic maskers, led to high between-language agreement. The effectiveness of the competing speech masker relative to SMN showed more L1-dependence.

It should be noted that our findings are based on short tokens of speech which are not in any way representative of natural sentences. For longer segments of speech we would expect to see far less language-independence due both to reduced experience of the target language's suprasegmental structure, word-boundary cues, lexical knowledge and syntax, as well as increasing L1 interference from these factors. However, what the current study demonstrates is that the low-level substrate on which these later processes are built itself shows relatively little L1 influence.

4.2. Effects of noise type

The ranking of noise conditions in the current study is consistent with the findings from earlier studies comparing modulated and unmodulated maskers (Festen and Plomp, 1990; García Lecumberri and Cooke, 2006) which demonstrated a clear reduction in masker effectiveness for modulated maskers, probably due to the possibility of glimpsing fragments of speech in the epochs of favourable SNR (Cooke, 2006; Fullgrabe et al., 2006).

The proportion of transmitted information for each feature varied with noise type, as did the relative rankings of feature values, in agreement with Jiang et al. (2006), who demonstrated that the ranking of perceptual cues changes in noise. However, the phonetic feature analysis led to some unexpected results. Voice was found to be the least well-transmitted feature in all conditions and for all listener groups. This contradicts previous studies which found voice to be best transmitted in noise (Miller and Nicely, 1955; Wang and Bilger, 1973). In the current study, place was much better transmitted than voice for all listener groups. Most earlier studies used fewer consonants, lacked certain phonological voicing contrasts and in most cases the only sonorants included were nasals (Miller and Nicely, 1955; Hazan and Simpson, 2000; Benki, 2003). The consonant set used here includes all obstruent voicing contrasts and the full set of sonorants. However, an analysis of identification scores for individual consonants suggested that the extra contrasts and sonorants cannot account for the voicing discrepancy. Instead, it is intriguing to note that a recent repeat of the Miller and Nicely experiment (Lovitt and Allen, 2006) found that voice was poorly transmitted, in both low and high noise conditions, and worse than place. It is possible that methodological differences such as live versus prerecorded stimulus presentation might account for observed differences between the studies.

Sibilants were robust in SSN, agreeing with previous studies (Miller and Nicely, 1955; Wang and Bilger, 1973; Wright, 2004). Sibilants are less vulnerable to energetic masking because they have much more intense noise components than other fricatives. Additionally, their high frequency energy allows them to escape some of the masking effect of SSN, whose spectrum falls with frequency.

4.3. The effect of an intelligible masker

Four of the seven non-native listener groups were more adversely affected by the competing speech masker than by a modulated noise masker. Since both maskers were constructed to produce similar amounts of energetic masking, it is tempting to suggest that this finding relates to differential informational masking effects between L1 groups. However, in practice it is not possible to achieve identical EM for this pair of maskers. Rather, while CS and SMN produce a near-identical temporal masking pattern, their spectral masking effect is identical only in terms of the long-term average. In particular, the CS masker has spectral fine structure, including harmonic peaks and valleys, while SMN has a much smoother spectral envelope. As a result, observed differences in masking effectiveness between SMN and CS could be due to both traditional informational masking (e.g. the effect on attention and cognitive load) and to the ability of listeners to exploit differing types of glimpses of the target stimulus made available by the presence or absence of fine structure in the masker spectrum.

Phonetic feature analysis revealed that the relatively adverse effect of the CS masker was largely due to worse reception of place and, to some extent, sibilant information relative to SMN. Since place is mainly conveyed by spectral cues, this finding points to a difficulty in segregating target and masker information within the spectrum for a competing talker. Source segregation (Darwin, 2008) is assumed to require both low-level processes which exploit, for instance, differences in fundamental frequency between foreground and background sources, as well as higher level knowledge. Since native and non-native listeners are capable of exploiting F0 differences to a similar extent in separating simultaneous sentences (Cooke et al., 2008), it seems likely that top–down influences were responsible for the differing masking effectiveness of CS and SMN. Some listener groups, and especially the native group, may be better able to allocate audible glimpses of foreground and background to the correct speech source and perhaps to make use of the upcoming pitch pattern. Likewise, it is possible that this benefit extends to processing of the competing speech since it was in the same language as the masker.

There is certainly a suggestion in Fig. 3 that the additional masking effect of a competing speaker tends to affect those groups whose consonant identification scores are lowest: the four L1s showing a more adverse effect of CS are among the five lowest-scoring groups. Coupled with the moderate negative correlation between individual listener performance overall on the task and the adverse effect of competing speech, this supports the idea that individual differences in prior knowledge of the L2 which help consonant identification *per se* also help to resist the effect of a competing talker, at least when the masker uses the same L2.

The fact that the CS masker was in the L1 of the native group might be expected to produce a distracting impact for the English listeners relative to the other language groups. García Lecumberri and Cooke (2006) found that monolingual English listeners were more adversely affected by a competing speaker in their L1 than when the masker was in an unfamiliar language (Spanish). However, in that study Spanish learners of English showed equivalent performance in the two maskers, so it may be that familiarity with the masker language rather than nativeness is sufficient to produce a distracting effect. All our listeners knew English, so the distracting effect may have been somewhat similar for all participants. Further studies using both L1 and L2 target items crossed with L1 and L2 maskers are required to separate effects of language familiarity and nativeness. It is worth recalling that the maskers used here were of 1.2 s duration, and while this is sufficient for the background to be largely intelligible, it is possible that longer stimuli would produce a greater effect of distraction.

4.4. Response times

Across all listeners and all consonants, a weak correlation was observed between better performance and shorter response times for all test backgrounds. While an analysis of individual consonants is outside the scope of this paper, significant variation in RTs across consonants exists, with the four "slowest" consonants viz. /ð,dʒ,θ,ʒ/ having response times nearly half as long again as the fastest, /l,m,n,r/. Here, there is a clear correspondence between accuracy and fast responding, with the slower group averaging 58% compared to 78% for the faster set.

Slower RTs might be thought to be associated with increased processing difficulty more generally, in which case longer response times would be expected for the masker which resulted in worst overall performance (SSN) compared to the other noise types. However, there was no RT difference between SSN and the other noise types. Comparing six L1 groups (leaving out the English and Dutch), one might expect to see slower mean response times in noise, reflecting the wide variation in consonant identification performance across these groups. Instead, response times were statistically-equivalent, again confounding the expectation that response time signals difficulty.

4.5. Methodological issues

A large-scale multilingual study brings with it intrinsic difficulties. Homogeneity of listener groups is hard to achieve in all non-native studies due to the great number of individual variables which affect performance. This problem is compounded when populations are extracted from varied countries and L1 backgrounds. Sample variability can obscure significant inter-group similarity. It is therefore possible that the similarity of rankings observed in the current study *under-estimate* the real degree of concordance. Due to the importance of individual variables in non-native speech perception, it will be of interest to repeat the listening tests of the current study with additional listener samples from the languages used here, and future studies would benefit from the inclusion of other Indo-European languages, particularly from more distant branches, and also from other language families. Stimuli, software and instructions for running these listening tests are available from the authors.

Another issue which affects speech perception research is the choice of whether to use phonetically-trained or naïve listeners. The latter provide a more representative sample of the general population but some of their responses are warped by orthographic influences, which differ according to L1 background. Responses from the former group may be more reliable but less generalisable to other listeners. It is notable that there was no correlation between self-assessed competence and listener group performance.

5. Conclusions

This study investigated the effects of different masker types and stimulus-related variability on the identification of consonants by eight listener groups differing in their L1s. The purpose of this multilingual sample was to reveal the extent of perceptual processes which are language-independent rather than language-specific. This approach differs in emphasis from classical treatments of non-native speech perception which focus on the influence of the L1 on the perception of NN sound contrasts (Flege, 1995; Best, 1995; Kuhl, 1993).

Several outcomes supported the idea that many aspects of intervocalic consonant identification are largely independent of the native language of the talker. Strong between-language similarity was seen in the ranking of the proportion of information transmitted for phonetic features, and the way this value changes in different noise conditions. Similarly, individual speaker intelligibility and the effects of gender, syllable stress, vowel context and token onset time were remarkably similar across listener groups. These findings have implications for the design of general-purpose signal enhancements designed to improve intelligibility in noise in environments populated by people listening in a foreign language.

One key difference between listener groups was seen in the effect of a competing speaker. Some non-native listener

groups found competing speech in the same language as the target tokens more disruptive relative to a speech-modulated noise masker designed to produce similar amounts of energetic masking. The adverse impact of a competing speaker was smaller for listeners who performed well in the task overall, suggesting that the prior knowledge of the target language which helps in L2 consonant identification in noise also assists in dealing with a competing speaker.

Acknowledgements

Corpus recording, annotation and native English listening tests took place while Martin Cooke was at the University of Sheffield, UK. We extend our thanks to Francesco Cutugno, Mircea Giurgiu, Bernd Meyer and Jan Volin for coordinating listener groups in Naples, Cluj-Napoca, Oldenburg and Prague; Youyi Lu (University of Sheffield) for speech material; Stuart Rosen (UCL) for making available the FIX software package; and the developers of the R statistical language R Development Core Team (2008). All authors were supported by the EU Marie Curie Research Training Network "Sound to Sense". Odette Scharenborg was supported by a Veni-grant from the Netherlands Organisation for Scientific Research (NWO). We also thank Marc Swerts and the reviewers for their insightful comments on an earlier version of the paper.

References

Ainsworth, W., Meyer, G., 1994. Recognition of plosive syllables in noise: comparison of an auditory model with human performance. J. Acoust. Soc. Am. 96, 687–694.

Alamsaoutra, D.M., Kohnert, K.J., Munson, B., Reichle, J., 2006. Synthesized speech intelligibility among native speakers and non-native speakers of English. Augment. Altern. Commun. 22, 258–268.

Barker, J., Cooke, M., 2007. Modelling speaker intelligibility in noise. Speech Commun. 49, 402–417.

Benki, J., 2003. Analysis of English nonsense syllable recognition in noise. Phonetica 60, 129–157.

Best, C., 1995. A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. Timonium, pp. 171–204.

Bradlow, A., Alexander, J., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Am. 121, 2339–2349.

Bradlow, A., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Am. 112, 272–284.

Bradlow, A., Pisoni, D., 1999. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. J. Acoust. Soc. Am. 106, 2074–2085.

Bradlow, A., Torretta, G., Pisoni, D., 1996. Intelligibility of normal speech I: global and fine-grained acoustic–phonetic talker characteristics. Speech Commun. 20, 255–272.

Brungart, D., Simpson, B., Ericson, M., Scott, K., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 100, 2527–2538.

Carhart, R., Tillman, T., Greetis, E., 1969. Perceptual masking in multiple sound backgrounds. J. Acoust. Soc. Am. 45, 694–703.

Cervera, T., Ainsworth, W., 2005. Effects of preceding noise on the perception of voiced plosives. Acta Acust. 91, 132–144.

Cooke, M., 2006. A glimpsing model of speech perception in noise. J. Acoust. Soc. Am. 119, 1562–1573.

Cooke, M., Scharenborg, O., 2008. The interspeech 2008 consonant challenge. In: Proceedings of the Interspeech. pp. 1765–1768.

Cooke, M., García Lecumberri, M., Barker, J., 2008. The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. J. Acoust. Soc. Am. 123, 414–427.

Cutler, A., Norris, D., 1979. Monitoring sentence comprehension. In: Cooper, W., Walker, E. (Eds.), Sentence Processing. Erlbaum, pp. 171–204.

Cutler, A., Mehler, J., Norris, D., Segui, J., 1987. Phoneme identification and the lexicon. Cognit. Psychol. 19, 141–177.

Cutler, A., Weber, A., Smits, R., Cooper, N., 2004. Patterns of English phoneme confusions by native and non-native listeners. J. Acoust. Soc. Am. 116, 3668–3678.

Cutler, A., García Lecumberri, M., Cooke, M., 2008. Consonant identification in noise by native and non-native listeners: effects of local context. J. Acoust. Soc. Am. 124, 1264–1268.

Darwin, C., 2008. Listening to speech in the presence of other sounds. Philos. Trans. R. Soc. B 363, 1011–1021.

Detey, S., Nespoulous, J., 2008. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonant clusters. Lingua 118, 66–81.

Dubno, J., Levitt, H., 1981. Predicting consonant confusions from acoustic analysis. J. Acoust. Soc. Am. 69, 249–261.

Festen, J., Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J. Acoust. Soc. Am. 88, 1725–1736.

Flege, J., 1995. Second language speech learning: theory, findings and problems. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. Timonium, pp. 233–277.

Florentine, M., Buus, S., Scharf, B., Canevet, G., 1984. Speech reception thresholds in noise for native and non-native listeners. J. Acoust. Soc. Am. 75, s84.

Foss, D., Blank, M., 1980. Identifying the speech codes. Cognit. Psychol. 12, 1–31.

Fullgrabe, C., Berthommier, F., Lorenzi, C., 2006. Masking release for consonant features in temporally fluctuating background noise. Hear. Res. 211, 74–84.

Gamer, M., Lemon, J., Fellows, I., 2007. irr: Various Coefficients of Interrater Reliability and Agreement. R Package Version 0.70. <http://www.r-project.org>.

García Lecumberri, M.L., Cooke, M.P., 2006. Effect of masker type on native and non-native consonant perception in noise. J. Acoust. Soc. Am. 119, 2445–2454.

Hazan, V., Markham, D., 2004. Acoustic–phonetic correlates of talker intelligibility for adults and children. J. Acoust. Soc. Am. 116, 3108–3118.

Hazan, V., Simpson, A., 1998. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. Speech Commun. 24, 211–226.

Hazan, V., Simpson, A., 2000. The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. Lang. Speech 43, 273–294.

Imai, S., Walley, A., Flege, J., 2005. Lexical frequency and neighborhood density effects on the recognition of native and Spanish accented words by native English and Spanish listeners. J. Acoust. Soc. Am. 117, 896–907.

Jiang, J., Chen, M., Alwan, A., 2006. On the perception of voicing in syllable-initial plosives in noise. J. Acoust. Soc. Am. 119, 1092–1105.

Kendall, M., 1948. Rank Correlation Methods. Griffin.

Kuhl, P., 1993. An examination of the perceptual magnet effect. J. Acoust. Soc. Am. 93, 2423.

Loizou, P., 2007. Speech Enhancement: Theory and Practice. CRC Press.

Lovitt, A., Allen, J., 2006. 50 years late: repeating Miller–Nicely 1955. In: Proceedings of the Interspeech. pp. 2154–2157.

Lu, Y., 2010. Production and Perceptual Analysis of Speech Produced in Noise. Ph.D. Thesis. University of Sheffield.

Mackay, I., Meador, D., Flege, J., 2001. The identification of English consonants by native speakers of Italian. Phonetica 58, 103–125.

Mayo, L., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech Lang. Hear. Res. 40, 686–693.

Miller, G., Nicely, P., 1955. Analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27, 338–352.

Moore, B., 2004. An Introduction to the Psychology of Hearing. Academic Press.

Parikh, G., Loizou, P., 2005. The influence of noise on vowel and consonant cues. J. Acoust. Soc. Am. 118, 3874–3888.

Picheny, M., Durlach, N., Braida, L., 1985. Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech. J. Speech Hear. Res. 28, 96–103.

Pinheiro, J., Bates, D., 2000. Mixed-Effects Models in S and S-PLUS. Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2008. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-89.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. <http://www.r-project.org>.

Rogers, C., Lister, J., Febo, D., Besing, J., Abrams, H., 2006. Effects of bilingualism, noise and reverberation on speech perception by listeners with normal hearing. Appl. Psycholinguist. 27, 465–485.

Studebaker, G., 1985. A rationalized arcsine transform. J. Speech Hear. Res. 28, 455–462.

Takata, Y., Nabelek, A., 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. J. Acoust. Soc. Am. 88, 663–666.

Van Engen, K., Bradlow, A., 2007. Sentence recognition in native- and foreign-language multi-talker background noise. J. Acoust. Soc. Am. 121, 519–526.

van Wijngaarden, S., Steeneken, H., Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native listeners. J. Acoust. Soc. Am. 111, 1906–1916.

Wang, M., Bilger, R., 1973. Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Am. 54, 1248–1266.

Wright, R., 2004. A review of perceptual cues and cue robustness. In: Hayes, B., Kirchner, R., Steriade, D. (Eds.), Phonetically Based Phonology. Cambridge University Press, pp. 34–57.