The non-native consonant challenge for European languages

M. Luisa García Lecumberri[1], Martin Cooke[2], Francesco Cutugno[3], Mircea Giurgiu[4], Bernd T. Meyer[5], Odette Scharenborg[6], Wim van Dommelen[7], Jan Volin[8]

[1] Department of English Philology, University of the Basque Country, Spain
[2] Department of Computer Science, University of Sheffield, UK
[3] Department of Physics, University "Federico II", Naples, Italy
[4] Department of Telecommunications, Technical University of Cluj-Napoca, Romania
[5] Medical Physics Section, Institute of Physics, University of Oldenburg, Germany
[6] Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands
[7] Department of Language and Communication Studies, NTNU, Norway
[8] Institute of Phonetics, Charles University in Prague, Czech Republic
garcia.lecumberri@ehu.es

Abstract

This paper reports on a multilingual investigation into the effects of different masker types on native and non-native perception in a VCV consonant recognition task. Native listeners outperformed 7 other language groups, but all groups showed a similar ranking of maskers. Strong first language (L1) interference was observed, both from the sound system and from the L1 orthography. Universal acoustic-perceptual tendencies are also at work in both native and non-native sound identifications in noise. The effect of linguistic distance, however, was less clear: in large multilingual studies, listener variables may overpower other factors.

Index Terms: speech perception, non-native, noise, masking

1. Introduction

The Interspeech Consonant Challenge involves the identification of English intervocalic consonants presented in quiet and in a variety of noise conditions [1], a comparison which highlights the effect of *imperfect signals* on speech perception. The current study introduces the additional factor of *imperfect knowledge* by examining the performance of non-native listeners on the Consonant Challenge.

Non-native sound perception has been shown to be heavily influenced by the L1 sound system [2, 3, 4]. Phonetic distance [5], language competence [6], universal tendencies and orthography [7] have been mentioned as additional factors that influence non-native (NN) perception. Speech perception in noise is an everyday situation which native listeners (NLs) learn to cope with using native competence to compensate for masking. Non-native listeners (NNLs) find these conditions all the more challenging since they lack rich and robust categories and they are subject to L1 interference. The present study extends previous work to a multilingual level. Native English listener performance on the consonant challenge tasks is compared to NN listeners from seven different language backgrounds with different levels of competence and distance to the target language. Besides English (en), the language groups tested were Czech (cz), Dutch (du), German (ge), Italian (it), Norwegian (no), Romanian (ro) and Spanish (sp).

2. Perception Tests

Details of perception testing for the native listener group are described in [1]. A similar procedure was used for non-native groups. Each consonant was represented on a computer screen by its most logical and frequent grapheme combination in English with a sample word below with the sound in question highlighted (e.g. 'B' for 'Bee', 'J' for 'Jar, 'CH' for 'CHart'). The use of graphemes was considered necessary since we wanted to collect data from phonetically naive subjects. However, this choice represents a compromise between testing a normal population via spelling, which increases 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.

The NNL tests were carried out in 7 countries following the same presentation and instructions to testers. All tests were carried out in quiet labs or booths. Listeners were given explanations on the nature of the tasks and on the sound-grapheme correspondences. If a particular example word was felt to be confusing for a language group (due to cognates, for instance), a different example word was chosen.

All listeners filled in a brief questionnaire prior to starting the test. Information collected included their age and English competence level (on a 4 point scale from '1= basic' to '4= fluent'). They were also asked to report if they were aware of having any hearing problems. Experimenters were asked to describe the academic background of their listener groups and familiarity with phonetic symbols.

A total of 207 listeners participated in the experiment. Listeners who reported hearing problems (3), or were not natives of the language group (3) or did not complete all the conditions (9) were excluded. A subsequent analysis of age distributions for each language group revealed some significant variation. Figure 1 shows the overall age distribution. In order to reduce the variation, 13 listeners aged 40 or above were removed from the analysis. Even so, considering the multi-site and multi-linguistic nature of this study, listener variability presented a serious concern. Although most listeners were at university and were fairly proficient in English, it was impossible to completely control for factors such as academic background, task/symbol familiarity, quantity and type of foreign language (FL)

experience, exposure and motivation. Since self-reported competence level was found to be uncorrelated with performance, in order to compare between languages in a balanced manner it was necessary to introduce some homogeneity as far as phonetic competence was concerned. Thus, a further filtering of participants based on performance in quiet was performed. The non-native scores in quiet were sorted, after which a 5th order polynomial was fitted to the sorted scores (Figure 2). On the basis of the elbow at around 73%, 24 listeners with scores in quiet lower than this were removed. Table 1 provides a summary of the final population used in the analyses reported in this paper.

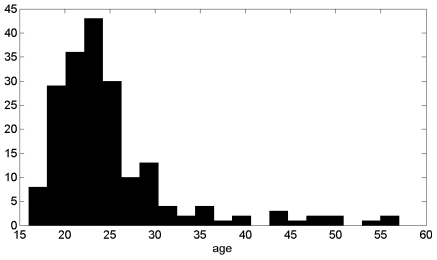

Figure 1 Raw age distribution (all languages)

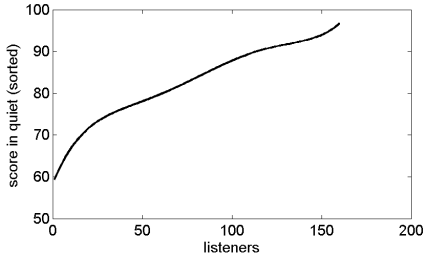

Figure 2. Non-native score distribution in quiet (5th order polynomial approximation).

Table 1. Listener group statistics. N (orig) indicates the final and original number of listeners prior to listener filtering.

	N (orig)	Age in years mean (sd), range	Proficiency [1-4]	Phonetic aware?	Test conditions
en	19 (25)	26.4 (5.5) 18-35	N/A	no	booth
cz	18 (18)	21.1 (2.9) 16-26	2.44 (0.62)	yes	booth
du	14 (23)	26.4 (5.0) 21-38	3.00 (0.68)	most no	lab
ge	19 (20)	26.4 (3.7) 21-33	2.42 (0.77)	most no	booth
sp	29 (52)	21.6 (3.6) 19-35	2.86 (0.44)	yes	lab
no	17 (21)	23.2 (3.0) 19-31	3.00 (0.61)	most no	lab
ro	25 (29)	22.8 (0.6) 21-24	2.68 (0.75)	no	lab
it	13 (20)	27.2 (3.1) 24-34	2.46 (0.66)	50% yes	lab

3. Results

3.1. Identification scores per language and condition

Table 2 shows the perception scores obtained by each listener group in each condition. Figure 3 displays the same information in a graph in which listening conditions are arranged in order of difficulty. As was expected, NLs showed better overall perception scores than any NNL group, although for Czech the difference was remarkably small. In terms of language distance, there was a tendency for listener groups from languages closer to English (German and Dutch) to be better than those from more distant languages (Romance group). As shown in Figure 3, all listener groups displayed a similar ranking of noise types.

Table 2. Mean consonant identification rates

	quiet	mean noise	talker	8-babble	SSN	factory	mod-babble	3-babble
testset	**1**		**2**	**3**	**4**	**5**	**6**	**7**
en	93.3	74.3	79.3	76.5	72.4	66.5	79.0	72.0
cz	92.9	72.0	75.5	74.2	70.5	65.4	77.2	69.3
ge	88.0	66.2	72.6	67.8	64.4	58.9	71.8	61.5
du	87.2	64.8	69.0	68.1	62.6	57.8	72.2	59.0
no	84.7	63.7	69.5	66.4	62.0	54.4	70.4	59.5
ro	83.8	60.4	61.8	62.0	59.4	54.7	67.3	57.2
it	82.2	59.4	64.5	62.1	54.7	53.6	66.8	54.5
sp	81.7	57.1	60.3	59.7	54.8	50.8	62.8	54.0

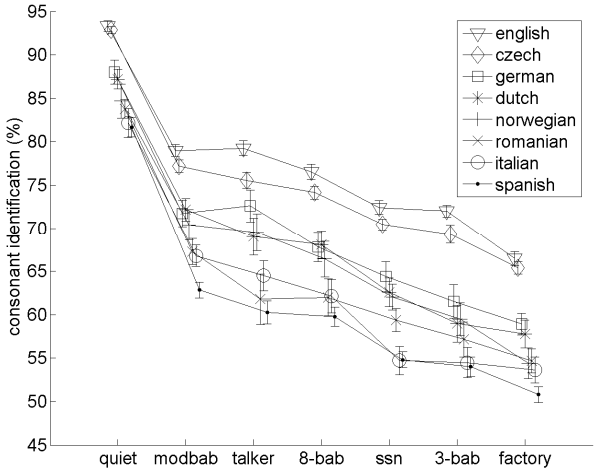

Figure 3 Mean identification scores (error bars represent +/- 1 standard error)

To determine the effect of noise on each listener group, the mean performance over the 6 noise conditions was computed. Figure 4 shows scores for quiet and the mean over noise as well as the proportion (noise/quiet). The ranking of listener groups in quiet was maintained in noise. Interestingly, the degree to which a group suffered in noise was inversely related to its performance in quiet. For example, in noise, natives scored around 79% of their performance in quiet, while for the Spanish group the equivalent figure was 70%. The proportional degradation in noise was statistically different for natives and non-natives as whole ($F=22$, $p < 0.001$). The difference was still significant for the three NN groups that carried out the experiment in the booth ($F=15.5$, $p < 0.001$), suggesting that differences in testing conditions cannot entirely explain the disproportionality.

3.2. Correlations

A correlation analysis was carried out between listener variables (Table 1) and performance (in quiet and mean over noise conditions). Self-reported proficiency was not correlated with perception in either quiet or noisy backgrounds, which points to the lack of reliability of self-reporting for phonetic research. Performance in quiet was a very good predictor of scores in noise ($r=0.825$, $p < 0.001$) and a fair predictor of the proportional degradation in noise (i.e. noise/quiet) ($r=0.33$, $p < 0.001$).

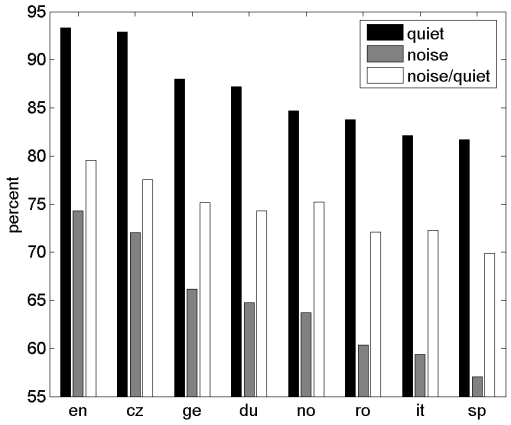

Figure 4. Scores in quiet, mean over noise conditions, and the proportion noise/quiet.

3.3. Consonant scores

Figure 5 shows the native advantage (i.e. the difference between native and non-native consonant identification rate in percentage points) in *quiet* for each language.

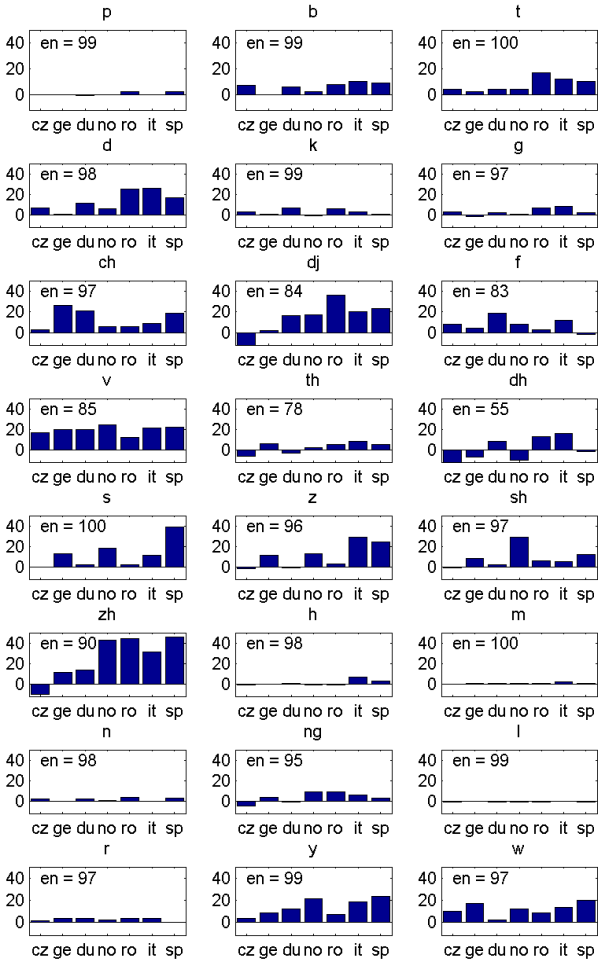

Figure 5. Native advantage over each language in quiet, expressed in percentage points. Baseline native performance is indicated in each panel (en). [‘ch’ = /tʃ/; ‘dj’ = /dʒ/; ‘th’ = /θ/; ‘dh’ = /ð/; ‘sh’ = /ʃ/; ‘zh’ = /ʒ/; ‘ng’ = /ŋ/; ‘y’ = /j/]

Consonants such as (/p k g h m n l r/) are well-identified by NLs and NNLs alike while others prove to be universally difficult (/θ ð f v dʒ/). Consonants such as /tʃ dʒ d v ʒ j w/ are problematic for most NNLs relative to NLs.

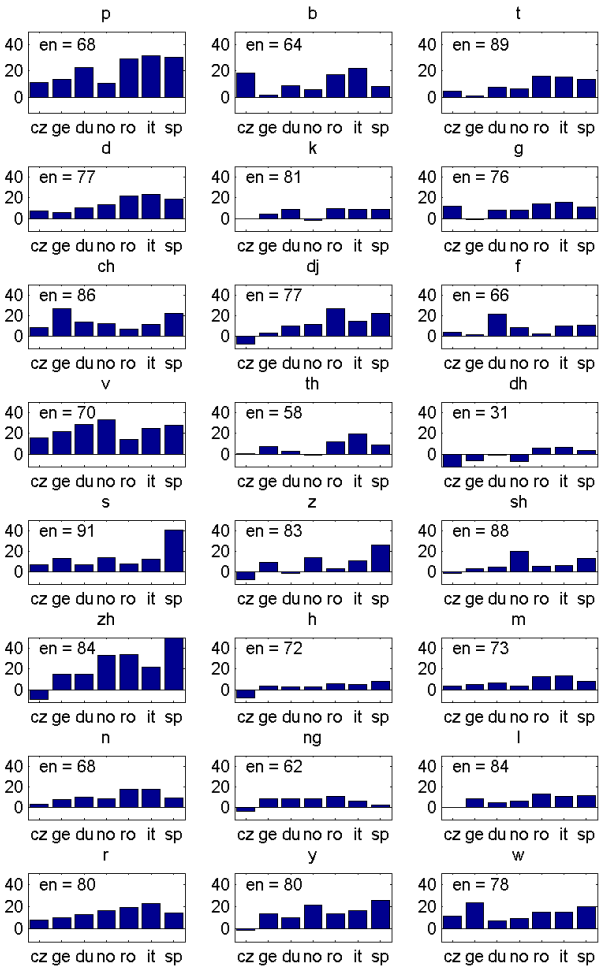

Figure 6. Native advantage in noise.

Most consonants which were difficult in quiet suffered further in noise and for NNLs often more so than for NLs. Figure 6 displays the native advantage in mean consonant identification rates across noise backgrounds for each consonant and language.

It is interesting to note that, in general, those consonants for which the NN disadvantage is largest in quiet do not suffer further disadvantage in noise. That is, for the sounds (/tʃ dʒ ʒ j w v d/), NLs and NNLs are equally affected by noise, albeit from a different baseline in quiet.

4. Discussion

Native/non-native comparisons are fraught with a great number of individual variables such as competence, motivation and exposure that can never be totally controlled, a problem which is magnified in multi-language studies such as the one presented here. Although the methodology and task were identical across listener groups, their backgrounds differed to a sufficient degree to bring in additional sources of variability: speakers varied in their English competence level as well as in their metalinguistic and phonetic knowledge. Thus, for some listeners, the use of orthographic sound representations was the only viable alternative whereas for

others it may have been a source of confusions. For instance, the letter 'g' often represents a voiced velar fricative in the languages of this study but in Italian and Romanian rather more often than in the others it corresponds to a voiced palatoalveolar affricate (before front vowels, eg. 'giro' (it) 'ger' (ro)), which shows up in the higher confusions these two language groups display for /g/ in quiet. Similarly, the native listeners in the present study had great difficulties with the sounds /θ ð/, largely due to spelling confusions, with such poor perceptions in quiet that several NN groups outperformed the natives, even though these two consonants were absent in some of their L1s (Norwegian, German, Czech and Spanish). The use of phonetic symbols would reduce the influence of orthographic confusions.

As expected, it was seen that NLs are better at consonant identification in quiet and noisy conditions. A detailed study of consonant scores and confusions revealed several tendencies. The group of consonants which displayed the worst perceptual results in quiet for NLs (/θ ð f v dʒ/) were found to be also amongst the worst for NNLs. These confusions have a clear acoustic-perceptual basis (/v ð/ /f θ/) or an orthographic motivation (/θ ð/). Dentals and labiodentals (which have lower RMS energy than sibilants) are very similar in their spectral characteristics of fricative noise, the main difference being in formant transitions.

The most robust sounds in noise are the sibilants (particularly voiceless) and /t/, whose high frequency burst resembles the sibilants' profile. This tendency is obscured in some language groups due to L1 interference rather than to masking (eg. /s/ is poor in the Spanish group and confused with /z/, whilst /tʃ/ is difficult for the Germans and confused with /dʒ/ because the second sound in each pair is mostly absent in the respective L1s and probably the object of overgeneralization [8] in the process of acquisition, and also because in both cases there is additionally a strong source of graphemic confusions).

On the other hand, one of the biggest NN deficits in noise corresponds to a sound with native-like perception in quiet (/p/). Interestingly, this is the sound for which there is the biggest NL drop from quiet to noise too, which indicates the presence of acoustic masking in noise which nevertheless affects NNs more, even though superficially the sound is 'equivalent' [3] to a category in all the NNs' L1s. Similarly, the nasals suffer a large perceptual deterioration for both NLs and NNLs – understandable since they are quite weak acoustically – but NNLs are less able to cope with masking despite the fact that they have near identical sounds (in the case of /n m/) in their L1s. Presumably, the disproportionate deterioration for NNLs is due to their inability to use perceptual cues or cue weightings which NLs draw on in adverse conditions. See also [9, 10, 11].

Other sounds with native-like perception in quiet (/l r/) suffer more in non-native noise but NLs are relatively unaffected. The disproportionate deteriorations of /l r/ are probably due to the different realizations of these two sounds (particularly /r/) in English in comparison to the other languages. Thus, although the English variant is easily recognizable in noise, it is a fragile category for NNs.

Although both NLs and NNLs found the different masking conditions similar in terms of their relative difficulty, NNLs suffered proportionally more in the presence of masking relative to their quiet scores. In a task such as the present one, which excludes the use of higher level information, this may be seen as an indication of either the fragility of their FL categories, of their lack of rich representations such as NLs possess which include cues which may be used in adverse conditions and/or of their use of cues or cue weighting different to those employed by NLs which may respond differently to masking.

5. Conclusions

A multilingual experiment comparing native and non-native intervocalic consonant perception in noise confirmed previous results indicating native advantage in all conditions as reflected in consonant perception scores and a disproportional deterioration of non-native scores in the presence of maskers. The latter may be explained by the NNLs' lack of robust and rich category representations available via the extensive and varied exposure typical of an L1. Perception in noise by both NLs and NNLs displayed acoustic-perceptual confusions due to masking and confusions due to orthographic interference. Similarly, sound robustness in noise could be appreciated across languages. Language background-related differences were found, such as a general tendency for linguistically closer languages to perform better, but these trends may be overshadowed by other inter-group differences. Nevertheless, there were clear perceptual patterns which could be explained by listeners' respective L1 interference, an acknowledged source of biases in non-native sound perception.

6. Acknowledgements

This work was carried out as part of the EU Marie Curie Research Training Network "Sound to Sense". We thank T. Brand for valuable input. B. Meyer is supported by DFG (SFB/TR 31 'The active auditory system'). O. Scharenborg is supported by a Veni-grant from NWO, The Netherlands.

7. References

[1] Cooke, M.P. and Scharenborg, O., "The Interspeech 2008 Consonant Challenge", submitted to Interspeech, 2008.

[2] Best, C. T., "A direct realist view of cross-language speech perception", in W. Strange [Ed], Speech Perception and Linguistic Experience, 171-204, Timonium, MD, 1995.

[3] Flege, J. E., "Second language speech learning: Theory, findings and problems", in W. Strange [Ed], Speech Perception and Linguistic Experience, 233-277, Timonium, MD, 1995.

[4] Kuhl, P.K., "An examination of the 'perceptual magnet' effect", J. Acoust. Soc. Am. 93, 2423, 1993.

[5] Hazan, V. and Simpson, A. "The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects ", Language and Speech 43: 273-294, 2000.

[6] Imai, S., Walley, A. C. and Flege, J. E., "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, 2005.

[7] Detey, S. and Nespoulous, J.L., "Can orthography influence second language syllabic segmentation?", Lingua, 118:66-81, 2008

[8] Major, R.C., "Phonological similarity, markedness, and rate of L2 acquisition", Studies in Second Language Acquisition, 9:63-82, 1987.

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

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

[11] Cutler, A., Cooke, M., Garcia Lecumberri, M. L. and Pasveer, D., "L2 consonant identification in noise: Cross-language comparisons", in Proc. INTERSPEECH 2007, Antwerp, 1585-1588, 2007.