L2 Consonant Identification in Noise: Cross-language Comparisons

Anne Cutler[1,2], *Martin Cooke*[3], *Maria Luisa Garcia Lecumberri*[4] *& Dennis Pasveer* [1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] MARCS Auditory Laboratories, University of Western Sydney, NSW, Australia
[3] Department of Computer Science, University of Sheffield, United Kingdom
[4] Department of English Philology, University of the Basque Country, Spain

anne.cutler@mpi.nl, m.cooke@dcs.shef.ac.uk, garcia.lecumberri@ehu.es, dennis.pasveer@gmail.com

Abstract

The difficulty of listening to speech in noise is exacerbated when the speech is in the listener's L2 rather than L1. In this study, Spanish and Dutch users of English as an L2 identified American English consonants in a constant intervocalic context. Their performance was compared with that of L1 (British English) listeners, under quiet conditions and when the speech was masked by speech from another talker or by noise. Masking affected performance more for the Spanish listeners than for the L1 listeners, but not for the Dutch listeners, whose performance was worse than the L1 case to about the same degree in all conditions. There were, however, large differences in the pattern of results across individual consonants, which were consistent with differences in how consonants are identified in the respective L1s.

Index Terms: speech perception, consonants, noise, L1, L2

1. Introduction

Listening to speech in a noisy environment is, notoriously, more difficult in a second language (L2) than in the native language (L1). This phenomenon, familiar to every user of more than one language, has been quite extensively studied and demonstrated under laboratory conditions [8,9,10,11]. An important issue is the extent to which the disproportionate difficulty of L2 compared with L1 perception in noise is due to inaccurate phoneme identification.

Thus it could be that listeners can draw on a wider range of cues to L1 phonemes, and are better able to detect the L1 cues from minimal evidence; in the L2, in contrast, listeners may be (as yet) unable to exploit the full range of cues, less able to identify those portions of the input most informative in noise [1], or only able to recognise the cues they use on the basis of clear and unambiguous evidence. If phonemes are incorrectly recognised, no word recognition is possible, and the whole speech perception process more or less falls apart.

On the other hand, it could be that the auditory perception accuracy of L1 and L2 listeners is equivalent, that for both L1 and L2 input, serious noise causes serious disruption of phoneme identification, but that the crucial difference lies in the degree to which recovery from the disruption is possible. In the L1, extensive experience with the distributional patterns of phonemes, of words, of phrases and of syntactic and semantic constructions pays off in realistic expectations of the probabilities for replacement of missing or poorly perceived portions of the input. In the L2, insufficient experience has been accrued for such realistic expectations to be rapidly derivable. The process of recovery from disruption is then too slow for speech perception to be repaired.

Of course, it is also possible for both scenarios to hold! There are indeed several findings indicating that L1 listeners make better use of contextual probabilities than L2 listeners [10,14], but also findings suggesting that L1 listeners make better use of phonetic cues [11]. Some studies have, therefore, attempted to focus as narrowly as possible on the phoneme perception process, in the hope of ascertaining whether it is indeed disproportionately disrupted by noise for L2 listeners.

Cutler, Weber, Smits and Cooper [4] presented American English vowels and consonants to L1 (American English) and L2 (Dutch) listeners, in three levels of masking by 6-talker babble. They found no evidence for disproportionate effects of noise on L2 listening. Noise affected L1 and L2 listeners to the same extent; L2 phoneme identification was about 80% of L1 performance across noise levels from 16 to 0dB SNR.

Garcia Lecumberri and Cooke [8] presented American English consonants to L1 (British English) and L2 (Spanish) listeners, either without masker or with one of several types of masker, including 8-talker babble. They found that the noise affected the L2 listeners more than the L1 listeners, such that L2 performance was 92% of L1 performance without masking but varied from 78% to 90% at 0dB SNR with the maskers.

The contrast in these patterns of results could have arisen from differences in the two experiments (enumerated in [8]) or from differences in the subject populations. The difference in babble talker N is unlikely to have had an effect since both Ns fell in the same performance range [13]. In both studies the stimuli were American English, but one L1 group was British, the other American, while one L2 group was Spanish and the other was Dutch. The L1 difference is unlikely to have played a role, given that Australian English listeners [5] matched American listeners' performance with the [4] stimuli. The L2 population difference might, however, have been crucial. Dutch and Spanish have very different phoneme repertoires. Dutch has (unusually among languages, especially those of Europe) a near-balanced repertoire of 19 consonants versus 16 vowels (13 monophthongs, three diphthongs), while Spanish has a highly unbalanced repertoire, quite typical of the world's languages: 20 consonants but only five vowels. This can have far-reaching consequences for vocabulary structure and hence for the task of word recognition [6,7].

Although the contribution of vowels versus consonants to word recognition is apparently equivalent in Spanish and Dutch [3], the difference in C:V ratio affects phoneme perception. In Dutch, effects of contextual (un)predictability are the same for vowel and for consonant identification, but in Spanish, effects of consonant context on vowel identification are greater than effects of vowel context on consonant identification [2]. This indicates that Spanish listeners are well aware of the greater potential for consonantal than for vocalic variation in their language, and that this awareness directly affects their phonemic decision-making. This difference cannot fully explain the different effect of noise on L2 listening observed in [4,8], since the different effects

were also found in [4,8] for consonant identification, which in [2] produced parallel performance from Dutch and Spanish listeners. What may be related to differences in performance across these groups, however, is inter-group differences in the type of speech information used for consonant identification, especially if the information in question is lost under noise masking; such inter-group differences have been attested [15].

In the present study we compare L2 listening performance of Spanish and Dutch listeners directly, on a single set of English materials. We confine our comparison to consonants only, using a new extended set which almost fully covers the repertoire of American English. We compare performance across consonant classes, and we compare these L2 listeners' performance with that of British listeners on the same stimuli.

2. Method

2.1. Participants

53 speakers of (European) Spanish, students in a phonetics class at the University of the Basque Country, 24 speakers of Dutch, students at the Radboud University Nijmegen, and 12 British English speakers, students at the University of Sheffield, took part in the study. None had any hearing impairment. The Spanish group received course credit, the other groups a small payment in return for participating.

2.2. Materials

These were the 23 consonants in the VCV materials constructed by Shannon et al. [12], in one vocalic context (α_α) only. The consonants were all those of American English with the exception of [ŋ] which cannot occur in this context. For each of the 23 consonants we used 10 tokens from separate talkers, for a total of 230 items. The materials were presented (a) against a quiet background without any masker, (b) with a masker of a single competing talker (American English, male speakers taken from the TIMIT corpus), or (c) with a masker of stationary speech-shaped noise generated by passing white noise through a filter set to respond with the long-term average spectrum of the same TIMIT speakers' productions). We refer to these conditions henceforth as "quiet", "talker" and "noise". In the talker and noise conditions, the SNR of the VCV was 0 dB; the masking started at VCV onset and continued until VCV offset.

2.3. Procedure

For all three groups, stimulus presentation and response collection was controlled by a computer running Matlab; the participants were informed that the test involved identification of English consonants, and they selected their response on each trial by clicking on one cell of a grid displaying letters or English words or phrases representing each of the 23 response options (for some consonants a word was available to provide the required intervocalic context – *alarm, afar* – but in most cases the context could best be provided in a phrase – *a yard, a bar*). Figure 1 shows the response grid used with the Dutch listeners; the grid was adjusted to suit the performance expectation for each listener group, with the British group clicking mainly on letters supplemented by a few words and the Spanish group using IPA symbols.

The stimuli were presented in quiet conditions over high-quality headphones (Plantronics for the Spanish group, Sennheiser for the British and Dutch groups).

a Path	a Bar	a Task	a Dart	a Car
a Garden	a CHart	aJar	aFar	a Van
a THought	oTHer	a Saw	baZaar	a SHark
a Genre	aHa	a Mark	baNana	aLarm
a Raft	a Yard	aWay		

Figure 1. Response panel for the experiment.

3. Results and Discussion

3.1. Overall

Figure 2 shows the mean percent correct consonant identification scores of each group in each condition. An overall repeated measures ANOVA revealed significant main effects of groups (F $(2,85)$=26.6, p < .001, η^2=0.38) and conditions (F$(2,85)$ p < .001, η^2=0.938). Within-condition analyses showed that all groups differed from one another in the quiet no-masker condition and in the stationary noise condition; in the single competing talker condition, each L2 group differed from the English L1 group, but the Spanish and Dutch group did not significantly differ.

Because the main question of interest concerns the degree of difference between each L2 group and the L1 comparison group, we further conducted separate ANOVAs for (English, Spanish) and (English, Dutch), with the conditions (quiet, competing talker) and (quiet, stationary noise) also treated separately within each language pair.
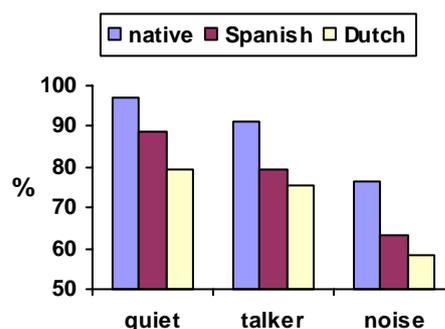

Figure 2. Percent correct consonant identification as a function of listener group (English native, Spanish L2, Dutch L2) and condition (quiet, talker, noise).

3.2. English vs. Spanish

We conducted separate analyses comparing each masking condition with quiet, looking for an interaction between listener group (L1, L2) and condition; this interaction will be significant wherever the effect of the masking is disproportionately greater for the L2 than for the L1 listeners.

The interaction between listener group and condition was significant in both cases: quiet vs. talker, p=0.001, η^2=0.162; quiet vs. noise, p=0.002, η^2=0.143. The Spanish listeners performed at 91% of the L1 scores in quiet, 87% with a competing talker, and at 83% in stationary noise. Thus the effect of both kinds of masker was clearly greater for the Spanish than for the L1 listeners, as also found in [8].

3.3. English vs. Dutch

In the case of the comparison of the Dutch L2 group with the L1 group, however, the interaction between listener group and condition was not significant for either condition: quiet vs. talker, p=0.31; quiet vs. noise p=0.69. Thus the effect of masking was clearly similar for the Dutch and for the L1 listeners, even though the overall performance of the Dutch listeners fell far below that of the L1 group, and even below that of the Spanish L2 group. The Dutch listeners performed at 82% of the L1 scores in quiet and with a competing talker, dropping to 76% in the stationary noise condition. In number of percentage points difference from the L1 performance, however, the Dutch (as Figure 3 shows) were reasonably uniform across the conditions, as also found in [4]. This was here in stark contrast to the Spanish listeners, whose scores in the talker and noise conditions were significantly further separated from the L1 performance than their scores in quiet.
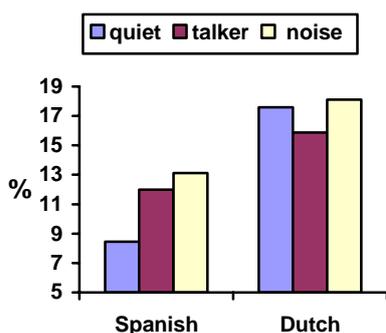

Figure 3: Non-native disadvantage per condition for the Spanish and Dutch group (percentage points difference from L1 performance)

3.4. Individual consonant scores

Figure 4 presents the difference between the L1 scores and each L2 listener group's scores for each of the 23 consonants. Differences for each condition (quiet, talker, noise) are in the leftmost three panels; the two rightmost panels present the additional difference between the disadvantage (L1-L2) in each masker condition versus the disadvantage in quiet.

It can immediately be seen that performance of the two L2 groups is not parallel, and further that for both groups, the consonants do not pattern uniformly across conditions. For some consonants, performance hardly differs from native L1 performance, while other consonants are much harder to identify for L2 than for L1. For instance, performance on the laterals and nasals in quiet is, for both L2 groups, hardly worse than L1 performance, while performance on the fricatives is, again for both groups, clearly worse than native performance in quiet. Likewise, the effects of each type of masking differ across consonants and across groups, and there are differences in the degree to which each masking type additionally increases the disadvantage that holds in quiet.
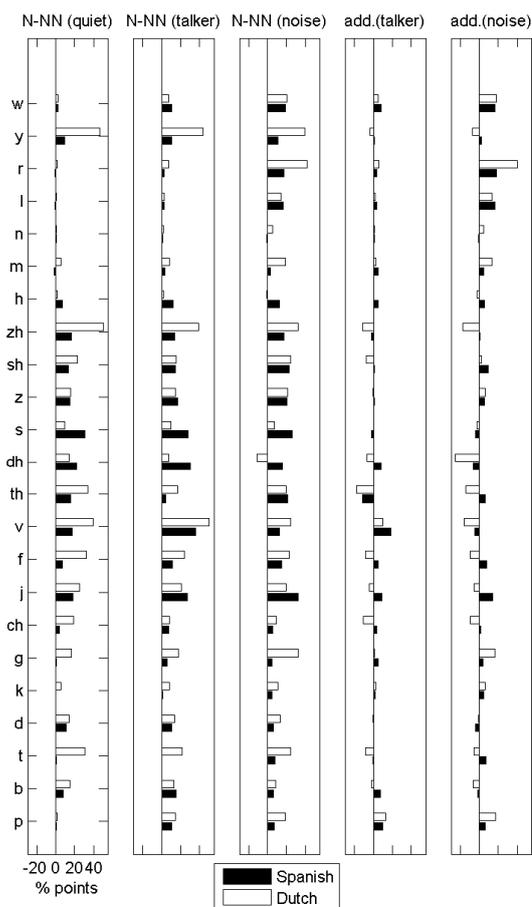

Figure 4: Non-native disadvantage per consonant for each L2 group, by condition

In this respect, the two rightmost panels provide information which sheds further light on the patterns displayed in the statistical analyses reported in sections 3.2 and 3.3. The overall pattern for the Spanish group is in agreement with the ANOVA outcome: the Spanish L2 listeners showed a small but consistent continuous decrement in performance from quiet to talker masker to noise masker. The interesting case is the Dutch group, because Figure 4 reveals a pattern concealed in the outcome of the L2 Dutch ANOVA. For some consonants, these L2 listeners' performance becomes even more distant from L1 levels with noise, but in many cases, noise actually reduces the difference between the performance of the Dutch group and the L1 group. The lack of an interaction between listener group and condition for the comparison of the L2 Dutch versus L1 English was therefore not at all consistent across the stimulus materials; the asymmetry with the L1 performance level increased for some consonants when the speech was masked, but actually decreased for others; averaged across all consonants, the asymmetry was thus zero.

The _decrease_ in these listeners' disadvantage compared with the L1 level, shown in the two rightmost panels, affects in particular the fricatives and affricates, and these phonemes are particularly affected in the stationary noise condition. We recomputed the analyses comparing masking versus quiet conditions for the Dutch L2 group versus the L1 group

(section 3.3), using the fricative and affricate data only. For this subset, the interaction between listener group and condition was significant both for quiet vs. talker, p=0.042, and for quiet vs. noise p=0.012. In contrast to the Spanish listeners, who showed a similar pattern with this subset of the data as overall (in quiet, 84% of L1 performance, deficit 15%; with competing talker 79%, deficit 18%; in stationary noise 73%, deficit again 18%), the pattern for the Dutch listeners was different from the overall pattern. With this subset they performed at 75% of the L1 scores in quiet but at 78% for each of the masking conditions. In number of percentage points difference from the L1 performance, the Dutch here closed the deficit from 24% in quiet to 19% and 15% in talker and noise respectively.

This is a very interesting finding in the light of recent research by Wagner et al. [15] on fricative identification by listeners with different L1s. They found that use of transition information for the identification of fricatives was not consistent across languages; listeners only used it where the L1 fricative inventory included confusable pairs which could more easily be distinguished if transitions were attended to. Thus both the English and Spanish fricative inventories contain the confusable pair /f,θ/; English and Spanish listeners attended to transition information in detecting occurrences of such phonemes. Listeners from languages without this pair did not attend to such information in detecting the same sounds. Among the languages in this latter group in Wagner et al.'s study was Dutch.

Translated to the present data set, this suggests that the two L2 groups tested here differed in the degree to which they used the cues which are used by L1 listeners for consonant identification. In the case of the fricatives, the similarities of native inventory structure led to similarities between the Spanish L2 listeners' cue use and the cue use of the English L1 listeners. The Dutch L2 listeners differed more from the English in respect to this subset of the data. However, the presence of a noise masker radically affected this pattern. Especially the stationary noise would make it difficult to use transitional cues, leading to a drop in performance for the L1 listeners and Spanish L2 listeners alike. Because the Dutch did not use these cues however, but relied on the steady-state information in the fricative noise [15], their relative performance drop from quiet to noise was less, which in turn resulted in the reduction in asymmetry between their performance and that of the L1 listeners for these items.

4. Conclusion

Listening to speech in noise is never easy, and it is harder when the speech in question is not in the L1. Our study focused on consonant identification in a constant context, a task in which listeners receive no support from knowledge of higher-level lexical or contextual information. We found that there were differences in how listeners from two language backgrounds performed in identifying American English consonants in noise. As previously observed [8], Spanish listeners' performance was more drastically affected by masking than was L1 listeners' performance. Also as previously observed [4], Dutch listeners' performance, though always worse than L1 performance, was not more affected by masking than L1 performance. The latter result, however, was in part due to different patterns of performance across different consonant types, which appear to result from cross-language differences in use of cues for identifying consonants.

6. References

[1] Cooke, M., "A glimpsing model of speech perception in noise", J. Acoust. Soc. Amer., Vol. 199, 2006, pp. 1562-1573.

[2] Costa, A., Cutler, A., and Sebastián-Gallés, N., "Effects of phoneme repertoire on phoneme decision", Perception & Psychophysics, Vol. 60, 1998, pp. 1022-1031.

[3] Cutler, A., Sebastián-Gallés, N., Soler Vilageliu, O., and van Ooijen, B., "Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons", Mem. & Cogn., Vol. 28, 2000, pp. 746-755.

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

[5] Cutler, A., Smits, R., and Cooper, N., "Vowel perception: Effects of non-native language versus non-native dialect", Speech Comm., Vol. 47, 2005, pp. 32-42.

[6] Cutler, A., Norris, D., and Sebastián-Gallés, N., "Phonemic repertoire and similarity within the vocabulary" Proc. Eighth Intern. Conference on Spoken Language Processing, Jeju, Korea, 1: 65-68, 2004.

[7] Cutler, A. and Pasveer, D., "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition", Proc. Speech Prosody 2006, Dresden, 237-400, 2006.

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

[9] Gat, I.B. and Keith, R.W., "An effect of linguistic experience", Audiology, Vol. 17, 1978, pp. 339-345.

[10] Mayo, L.H., Florentine, M., and Buus, S., "Age of second-language acquisition and perception of speech in noise", J. Speech Hear. Res., Vol. 40, 1997, pp. 686-693.

[11] Nabelek, A.K. and Donahue, A.M., "Perception of consonants in reverberation by native and non-native listeners", J. Acoust. Soc. Amer., Vol. 75, 1984, pp. 632-634.

[12] Shannon, R.V., Jensvold, A., Padilla, M., Robert, M.E., and Wang, X., "Consonant recordings for speech testing", J. Acoust. Soc. Amer., Vol. 106, 1999, pp. L71-L74.

[13] Simpson, S.A. and Cooke, M., "Consonant identification in N-talker babble is a nonmonotonic function of N", J. Acoust. Soc. Amer., Vol. 118, 2005, pp. 2775-2778.

[14] Van Wijngaarden, S., Steeneken, H., and Houtgast, T., "Quantifying the intelligibility of speech in noise for non-native listeners", J. Acoust. Soc. Amer., Vol. 11, 2002, pp. 1906-1916.

[15] Wagner, A., Ernestus, M., and Cutler, A., "Formant transitions in fricative identification: The role of native fricative inventory", J. Acoust. Soc. Amer., Vol. 120, 2006, pp. 2267-2277.