Effects of exposure to noise during perceptual training of non-native language sounds

Martin Cooke[1, a)] and Maria Luisa Garcia Lecumberri[2]

[1] *Ikerbasque (Basque Science Foundation)*

[2] *Language and Speech Laboratory, Universidad del País Vasco, 01006 Vitoria, Spain*

(Dated: 9 April 2018)

The following article appeared in **J. Acoust. Soc. Am.** 143, 2602–2610 (2018) and may be found at https://doi.org/10.1121/1.5035080

Listeners manage to acquire the sounds of their native language in spite of experiencing a range of acoustic conditions during acquisition, including the presence of noise. Is the same true for non-native sound acquisition? This study investigates whether the presence of masking noise during consonant training is a barrier to improvement, or, conversely, whether noise can be beneficial. Spanish learners identified English consonants with and without noise, before and after undergoing one of four extensive training regimes in which they were exposed to either consonants or vowels in the presence or absence of speech-shaped noise. The consonant-trained cohorts showed substantially larger gains than the vowel-trained groups, regardless of whether they were trained in noise or quiet. A small matched-condition benefit was evident, with noise-training resulting in larger improvements when testing in noise, and vice versa for training in quiet. No evidence for habituation to noise was observed: the cohort trained on vowels in noise showed no transference to consonants in noise. These findings demonstrate that noise exposure does not impede the acquisition of second language sounds.

---

[a] m.cooke@ikerbasque.org

I. INTRODUCTION

Acquiring the sounds of a first language is typically achieved in uncontrolled and at times noisy settings. In contrast, most formal training in the acquisition of a foreign language occurs in quieter conditions with fewer sources of interference than found in natural environments. Since the value of increasing input diversity has been demonstrated by high variability training regimes (Clopper and Pisoni, 2004; Logan *et al.*, 1991), it is natural to ask whether exposing language learners to noise might also be beneficial.

Noise is a real problem in non-native listening. While all listeners suffer in adverse noise conditions, non-native listeners are significantly challenged and can exhibit a disproportionate fall in intelligibility (Florentine *et al.*, 1984; García Lecumberri and Cooke, 2006; Takata and Nabelek, 1990); for a review, see García Lecumberri *et al.* (2010). While some of the native listener advantage in noise comes from their superior native language knowledge, it remains even in tasks such as consonant identification in vowel-consonant-vowel (VCV) tokens where semantic, syntactic and lexical information is not available, as long as some contextual information exists for native listeners to exploit (Cutler *et al.*, 2008).

There are a number of ways in which the presence of noise during the acquisition of non-native categories might be expected to benefit learners. One is by helping in the formation of robust sound categories. Non-native listeners are known to use cues and cue-weightings different from those used by native listeners (e.g., Bohn and Flege, 1990; Cebrian, 2006). Noise-based training might highlight those cues that are more resistant to masking (Lovitt and Allen, 2006; Miller and Nicely, 1955; Van Dommelen and Hazan, 2010; Wright, 2004), helping to weight their value in adverse conditions (c.f. weighting of speech segmentation cues in noise; Mattys *et al.*, 2005).

Another possibility is that listeners form exemplars which contain traces of both speech and noise, as suggested by studies with native listeners (Cooper *et al.*, 2015; Creel *et al.*, 2012; Pufahl and Samuel, 2014). This stance is analogous to the so-called 'multi-style' training shown to be effective in robust automatic speech recognition (e.g., Lippmann *et al.*, 1987). Alternatively, listeners who hear speech tokens in noise may learn to better handle the masker, or become more adept at the speech-in-noise task. Task effects could arise as a form of procedural learning (Koziol and Budding, 2012; Robinson and Summerfield, 2006) in which learners become familiarised with the properties of the masker (Wilson *et al.*, 2003). Alternatively, listeners might learn to tune out the masker through improved attentional focus.

On the other hand, training in noise might lead to a decrease in intelligibility. One effect of masking is to partially or completely obscure speech cues, so the quantity of useful speech information received during training can be expected to be lower than would be the case in the absence of noise. Noise may also increase attentional load, leading to fatigue or a reduction in resources available to process the incoming signal. It is therefore an open question as to whether masked presentation of tokens is an effective strategy for training non-native learners.

Speech in noise training has been explored in the past with native listeners, mainly for older adults with hearing deficits (e.g., Burk *et al.*, 2006; Humes *et al.*, 2009; Oba *et al.*, 2011; Stecker *et al.*, 2006; Woods *et al.*, 2015). The mean participant age in these studies ranged from 66.0 to 72.8 years. Most studies used words as training tokens. Training with words in noise has been shown to improve perception of trained tokens with the same or novel voices, but with limited generalisation to new materials or listening conditions. Indeed, Humes *et al.* (2009) argue that lack of generalisation to new words is due to the fact that training in noise is mainly a lexical process which helps to re-establish connections between the impoverished input and listeners' phonological representations in the lexicon. However, when using a closed set of digits in babble noise, Oba *et al.* (2011) found that improvements did generalise to another noise background and to other sentence materials.

The benefit of training in noise using nonsense syllables has also been found to generalise to other token types. Stecker *et al.* (2006) trained hearing impaired listeners on CV and VC nonsense tokens and obtained continuous improvements over an extensive number of training sessions. Initial gains were attributed to procedural learning (Robinson and Summerfield, 2006), but the fact that subsequent improvements extended to untrained voices and were retained in later post-testing was considered to be an indication of perceptual learning. In a similar vein, Woods *et al.* (2015) found substantial training benefits in listeners with mild to moderate hearing loss for consonant identification in noise in CVC syllables, with generalisation to novel speakers. While rapid initial gains were considered to be the result of procedural learning, improvements continued throughout the later stages of training. The authors ascribe these benefits to the use of a large corpus of varied stimuli, presented over a considerable period of time, and argue that the approach promotes perceptual learning.

A study with young normal hearing adults (mean age: 24.7) by Song *et al.* (2012) measured the effects of training in noise on two standard speech-in-noise tests (Killion *et al.*, 2004; Nilsson *et al.*, 1994), employing a sequence of 20 training sessions, each of 30 minutes duration. Training involved a range of adverse conditions including fast speech, simultaneous tasks, and two masking noise conditions where listeners heard speech in a multitalker babble or competing speech background. Relative to a control group, listeners improved significantly after training. Of relevance to the current study, Song *et al.* (2012) used a mixed cohort of native and non-native listeners, but unfortunately the results for the non-native group are not presented separately. As far as we are aware, there have been no studies of noise-based acquisition specifically focusing on non-native listeners.

The absence of data on the effect of noise exposure during second language acquisition motivates the current study, as a means to explore the wider issue of whether there are beneficial effects of acquiring speech sounds in less-than-pristine acoustic conditions. We address the question of whether exposing non-native listeners to noise during an extensive training period is an effective strategy for acquiring the consonants of a second language. Our design also allows us to determine whether learners are able to transfer any benefits of noise exposure to an untrained type of masker or speech token type.

In the current study, four homogeneous cohorts of Spanish learners of English underwent one of four training regimes, bracketed by an identical pre-test and post-test involving forced-choice identification of consonants in quiet, in speech-shaped noise, and in a babble masker. During 10 training sessions, two of the groups undertook forced-choice consonant identification in VCV tokens with feedback on incorrect responses. One of these groups performed the task without noise, while the other heard the same tokens mixed with a speech-shaped noise masker. Two further groups identified vowels in CVC tokens, one group in quiet, the other with noise. The vowel-trained groups served as controls, allowing an estimate of the effect of external factors such as concurrent exposure to English from other sources, or the effect of task familiarity. Comparison between the two vowel groups enables any noise-exposure transfer effect to be quantified. The use of an untrained masker (babble) also reveals any transfer of noise-training benefits to a novel masker.

In summary, this study tests the following hypotheses:

(i) Speech-in-noise training is an effective strategy for non-native consonant acquisition. This would be substantiated by a finding that the group trained on consonants in noise exhibits greater pre-to-post test gains than the groups trained on vowels. Additionally, comparing any gains with those of the group trained on consonants in quiet serves to quantify the degree of effectiveness of noise-based training.

(ii) Habituation to the presence of noise is responsible for some of the beneficial effects of noise-based training. This hypothesis would be supported if gains for consonants for the group trained on vowels in noise are seen to exceed those of the group trained on vowels in quiet.

(iii) Noise helps via the formation of robust cues or cue-weightings. This notion would be supported by finding any transfer of benefit to either the quiet or un-trained babble masker condition for the noise-trained consonant group.

II. METHODS

A. Listeners

A group of 88 native Spanish listeners (67 female; mean age 19.5 years, std. dev. 2.3) in the second year of study on a degree in English Philology at the University of the Basque Country took part in the experiment in return for course credit. Participants were either Spanish monolinguals or Spanish/Basque bilinguals. Apart from the presence in Basque of a palato-alveolar fricative akin to English /ʃ/, there are no relevant differences between Basque and Spanish for consonants in intervocalic positions. Listeners reported no hearing problems. In parallel with the training procedure, participants pursued a module in English Phonetics which included practice in the analysis and transcription of English vowels and consonants. Participants were familiar with the International Phonetic Alphabet (IPA) symbols for vowels and consonants at the outset of the training procedure.

B. Speech materials

Training and test materials were drawn from an existing source of British English consonant data, the Consonant Challenge Corpus (Cooke *et al.*, 2010; Cooke and Scharenborg, 2008). A subset of the corpus consisting of nonsense VCV tokens spoken by 12 male and 12 female talkers was selected for use in the current study. The subset contains 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 /iː, uː, æ/ for both front and end stress (e.g., /ˈæbiː/ versus /æˈbiː/), leading to a possible 10368 tokens. VCVs used in the testing phases came from four male and four female talkers, while those employed during training were derived from the remaining eight male and eight female talkers. VCVs ranged in duration from 290-1002 ms, with a mean duration of 602 ms.

Speech material used during the training phase for the vowel-trained groups consisted of monosyllabic CVC words (e.g.,“look”, “hid”, “sup”) spoken by 7 British English talkers. Each word contained one of 11 English vowels / iː, ɪ, e, æ, ʌ, ɑː, ɒ, ɔː, ɜː, ʊ, uː/.

C. Maskers

Two maskers were used in the current study. During the training phase, listeners in noise-trained groups heard tokens mixed with speech-shaped noise (SSN). In the pre- and post-tests, listeners in all experimental groups identified consonants masked by SSN and by an 8-talker babble masker (BAB) in separate condition blocks. Noisy tokens were generated by mixing speech with randomly-chosen masker fragments of 1.2 s duration. The onset of the speech relative to the noise was varied, taking on a value in the range 0-400 ms. The masker was scaled to produce the target signal-to-noise ratio (SNR) in the region containing the speech signal i.e., discounting the leading and lagging noise-only sections of the waveform. The noisy test sets correspond to test sets 3 (BAB) and 4 (SSN) of Cooke and Scharenborg (2008).

D. Consonant identification: pre- and post-tests

During the pre- and post-tests, which were identical in all respects, listeners first identified VCVs in quiet, followed by VCVs mixed with SSN at a token-wise SNR of -6 dB, and subsequently VCVs mixed with babble at a token-wise SNR of -2 dB. These SNR values were chosen in Cooke and Scharenborg (2008) to produce identification rates of around 70% for native listeners. Note that throughout the paper we refer to the three conditions as 'masking conditions' even though in the quiet condition the masker is absent.

In each of the three blocks listeners undertook a 24-alternative forced choice identification task under computer control by selecting a consonant from an onscreen keyboard containing IPA symbols for each consonant. Sixteen examples of each of the 24 consonants were used in each test block, made up of a front-stressed and an end-stressed exemplar from each of the eight talkers, leading to a total of 384 stimuli per block, some 1152 tokens across the three test blocks. All stimuli were distinct, with vowel contexts chosen at random. To familiarise themselves with the upcoming masker condition, listeners underwent a short practice session containing 16 stimuli prior to each of the two blocks containing noisy tokens. On average listeners required approximately 18 minutes to complete each block in the pre-test and 14 minutes for the post-test.

E. Assignment to experimental groups

Following the pre-test, listeners were assigned to one of four experimental groups. The CONS-Q group were trained on consonants in quiet, while the CONS-N group heard the same tokens mixed with the SSN masker. Similarly, the VOW-Q and VOW-N cohorts were trained on vowels in quiet and noise respectively. Twenty-two participants were assigned pseudo-randomly to each of the four groups following a group score balancing procedure in such a way as to satisfy the criterion that the four group mean scores were within 1 percentage point of each other in each of the three pre-test conditions.

F. Training procedure

All groups received perceptual training during 10 separate sessions over the course of 5 consecutive weeks. Training began in the week following the pre-test, and ended the week preceding the post-test. Each training session consisted of five equal-length blocks.

Listeners belonging to the CONS-Q and CONS-N groups identified four VCV tokens for each of the 24 English consonants in each block, i.e., 20 exemplars per consonant per session. The procedure was identical to the test phases except that listeners received feedback on incorrect responses and had to listen exactly once again to the stimulus before moving on to the next token. For the CONS-N group, each of the five blocks per session was presented at one of five SNRs: -2, 0, -2, -4 and -6 dB. Note that the most adverse SNR corresponded to that of the test phase, and the remaining SNRs were somewhat more favourable. A range of SNR values was chosen in order to promote variability in the availability of speech cues following masking, corresponding to acquisition in everyday noisy environments. Across the 10 training sessions listeners responded to a total of 4800 distinct tokens, 200 per consonant.

The two vowel groups also heard five blocks of vowel stimuli per session. Within each block, vowels came from the same talker. No talker was repeated in any individual session. Listeners received feedback as for the consonant-trained groups. Stimuli for the VOW-N group consisted of vowels mixed with SSN at an SNR of -6 dB. This value was chosen to match to the SNR used in the consonant test material.

All training sessions took place in a quiet language laboratory. Listeners heard stimuli through Plantronics Audio-90 headphones at a comfortable listening level that they were able to set individually.

G. Post-processing

Of the 88 participants, one member of the VOW-N group did not complete the training sessions and was excluded from the analysis. Another member of the VOW-N group showed a drop of 25 percentage points in one masked condition in the post-test relative to the pre-test, and was also removed from further analysis.

Listener performance was measured as the percentage of consonants identified correctly in each condition. Percentage correct scores were transformed to rationalised arcsine units (RAUs; Studebaker, 1985) for statistical testing. Since statistical outcomes with RAU scores were identical to those based on raw percentages, for ease of interpretation raw percentages are used in the text and in the results figures.

III. RESULTS

A. Consonant identification

Figure 1 depicts the percentage of correctly-identified consonants as a function of experimental group and test phase. Since the four experimental groups were assigned in such a way as to equate group mean scores for each of the three masking conditions, a single mean per condition is shown for the pre-test. Also shown for comparison are identification rates based on precisely the same speech-in-noise stimuli for the native English listener sample tested by Cooke and Scharenborg (2008). At the pre-test stage, non-native listener accuracy is 85% of that of natives in quiet (79.7% versus 93.8%) while for the masked conditions the equivalent figures are 79% for BAB (60.8% versus 76.5%) and 75% for SSN (54.1% versus 72.2%). All four groups showed an improvement by the time of the post-test, with gains ranging from 2.3 to 14.1 percentage points. To put these changes into perspective, the highest scoring group in quiet reached over 98% of the native score, while in BAB and SSN the highest-scoring groups obtained 94% and 95% of native performance. These figures attest to the impact of the training period, and suggest limited room for further improvement given a longer period of exposure (see also section III B below).

An analysis of variance (ANOVA) of RAU-transformed scores with within-subjects factors of masker type (quiet, SSN, BAB) and test time (pre, post), with experimental group as a between-subjects factor, indicated significant interactions between the three factors [$F(6, 164) = 4.8, p < .001, \eta^2 = 0.007$], between masker type and test time [$F(2, 164) = 21.5, p < .001, \eta^2 = 0.01$] and between group and test time [$F(3, 82) = 62.6, p < .001, \eta^2 = 0.11$], alongside significant main effects of group [$F(3, 82) = 4.83, p < .001, \eta^2 = 0.12$], masker type [$F(2, 164) = 2441, p < .001, \eta^2 = 0.76$] and test time [$F(1, 82) = 583, p < .001, \eta^2 = 0.29$]. These outcomes are explored in more detail below.

1. *Vowel-trained groups*

Gains for the vowel-trained groups allow for a quantification of any effects other than specific consonant training (for instance, gains due to procedural learning, exposure to noisy

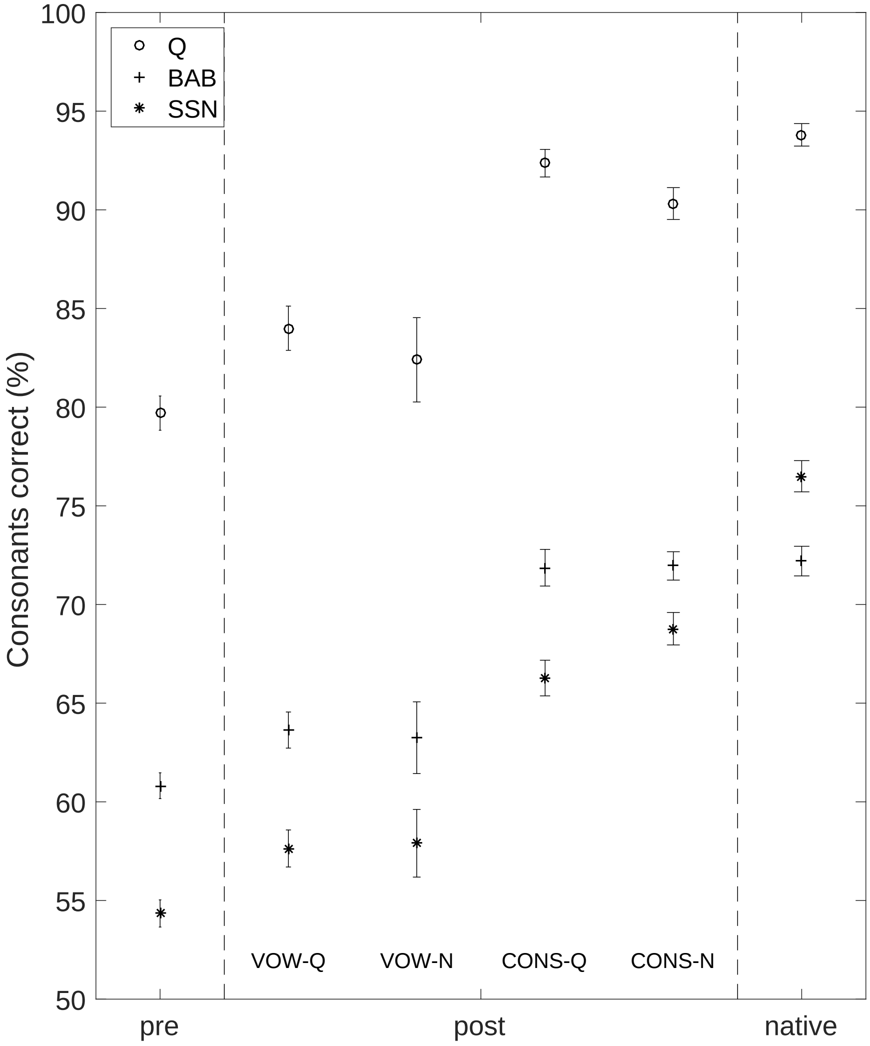

FIG. 1. Consonant identification rates. Column ‘pre’ denotes the mean score across all four groups in the pre-test while ‘native’ shows scores for native listeners taken from Cooke and Scharenborg (2008). The remaining columns correspond to the four experimental groups in the post-test. Error bars here and in subsequent figures denote ±1 standard error.

tokens during the pre-test or familiarisation with IPA symbols for response categories). Across noise conditions, gains ranged from 2.2 to 4.3 percentage points. Post-test scores were significantly higher than in the pre-test [$F(1, 40) = 10.00, p < .001, \eta^2 = 0.05$], with the smallest gain of 2.2 in the BAB condition for the VOW-N group exceeding a Fisher's Least Significant Difference (FLSD) of 1.2. However, there was no evidence of a transfer of benefits from exposure to noise during training from vowels to consonants. The two vowel groups did not differ in their post-test scores in any of the masker conditions, with no significant effect of group [$p = 0.86$] and no interaction with masker type [$p = 0.57$].

2. Consonant-trained groups

A clear effect of explicit consonant training is evident in the results: groups trained on consonants made substantially larger gains than the vowel-trained groups [$p(1, 84) = 63.5, p < .001; \eta^2 = 0.39$] overall. Consonant-trained groups out-performed vowel-trained groups by 8.1, 8.5 and 9.8 percentage points in the quiet, BAB and SSN conditions respectively, relative to a FLSD of 1.00 percentage point.

Considering the two consonant-trained groups, a two-factor ANOVA on RAU-transformed post-test scores with a between-subjects factor of group (quiet vs. noise training) and a within-subjects factor of masking condition revealed an interaction between group and masker [$F(2, 84) = 16.7, p < .001, \eta^2 = 0.06$] as well as the expected masking condition effect [$F(2, 84) = 1895, p < .001, \eta^2 = 0.89$]. The interaction is due to differences in the quiet and SSN conditions. The CONS-N group had higher scores than the CONS-Q cohort in the matched SSN condition (68.8% vs. 66.3%), a difference significantly larger than the FLSD

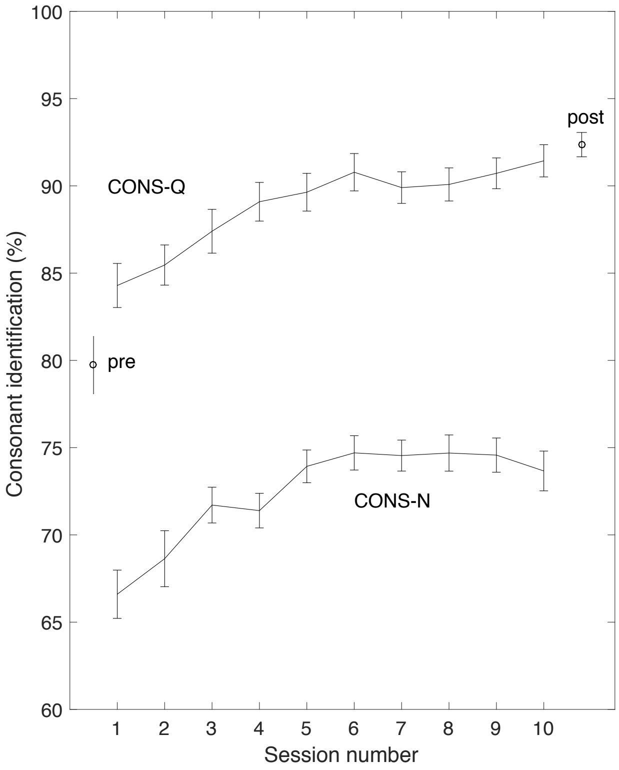

FIG. 2. Consonant identification rates in each training session for the quiet-trained (CONS-Q; listening in quiet) and noise-trained (CONS-N; listening in noise) groups. Identification rates in the quiet condition of the pre- and post-test for the quiet-trained group are also shown.

of 1.1. Conversely, the group trained in quiet identified a higher proportion of consonants in quiet compared to the noise-trained group (92.4% vs. 90.3%). Thus, each group showed a modest but statistically-significant matched-training benefit. In contrast, scores in the BAB condition were almost identical – 71.9% and 72.0% for the quiet and noise-trained groups respectively.

B. Evolution of consonant identification during training

Figure 2 depicts scores for the two consonant-trained groups during each of the 10 training sessions, along with the pre- and post-test scores for the CONS-Q group. Since the SNRs in test and training were not fully matched (see section II F) it is not meaningful to compare scores for the CONS-N group with their pre-test scores in the SSN masking condition. Of particular note is the difference of around four percentage points between the pre-test and initial training session of this group, which suggests that while no feedback was provided during training, familiarity with the task played a role in the initial improvement. Both cohorts exhibited a steady improvement over the first six sessions, with little or no improvement thereafter.

C. Identification rates and gains for individual consonants

Figure 3 displays mean identification scores in the pre-test for each consonant in the quiet and SSN conditions. Based on their location relative to the upper diagonal, which indicates equal scores in quiet and noise, and the lower diagonal, which denotes the mean reduction in noise, it is possible to identify three groups of consonants. One group consisting of the sibilants /ʃ, ʒ, z/) and the plosive /t/ shows no adverse effect of masking, most likely due to the quasi-low-pass spectrum of the speech-shaped masker which allows the intense high frequencies of sibilants and the aspiration noise of /t/ to escape masking (Hayward, 2002; Kent *et al.*, 1996; Kent and Read, 1992). Another group, notably /p, m, n, l, k/ and to a lesser extent /b, ŋ, f, h, g, r/, contains consonants that are well-identified in quiet but show

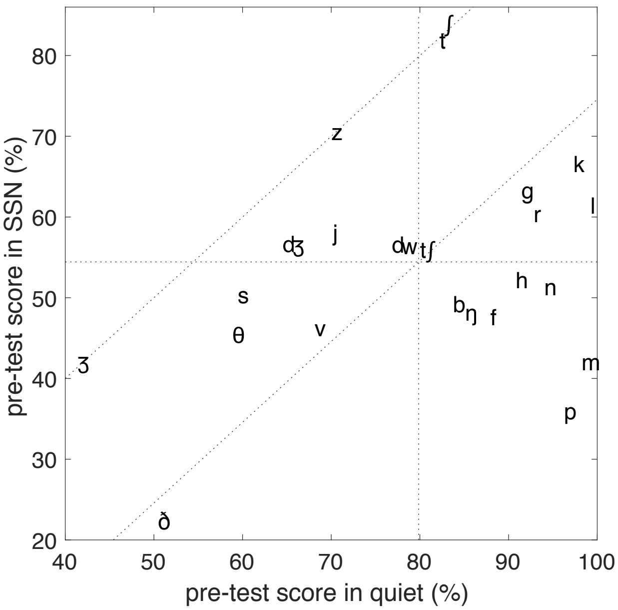

FIG. 3. Mean consonant scores in the quiet and SSN conditions of the pre-test. The vertical and horizontal lines indicate the mean identification rates in quiet and noise respectively. The upper diagonal line denotes equal identification scores in the two conditions, while the lower diagonal line separates consonants whose score reduction in noise lies above or below the average reduction.

above-average reductions in SSN. Most of the remaining consonants fall between these two extremes, with poor-to-moderate scores in quiet and small-to-moderate reductions in noise. The weak fricative /ð/ is something of an outlier, possibly because of the combined effects of low intensity and native language influences: orthographically, the equivalent sound in Spanish is written as “d”.

Figure 4 shows the changes in identification rates after training for each of the four experimental groups in the quiet and SSN testing conditions. Most sounds show gains in all four training groups although the improvements are generally much smaller for the two vowel-trained groups. Categories that were well-identified in the pre-test have reduced potential

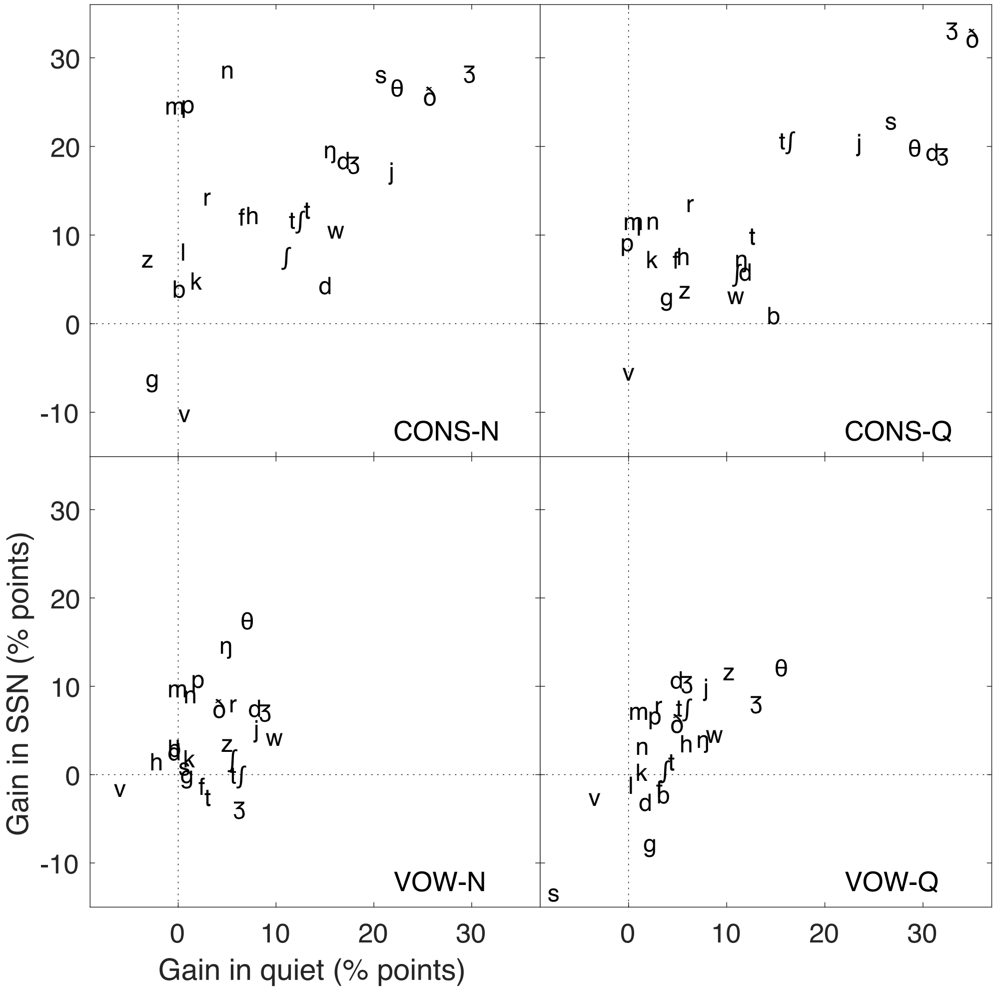

FIG. 4. Changes in consonant scores from pre- to post-test.

for further improvement in quiet. It is among the 8 consonants /z, j, v, dʒ, s, θ, ð, ʒ/ that have identification rates below 70% in the pre-test that we observe most of the substantial post-training gains for the CONS-Q group relative to the CONS-N group in the quiet testing condition. The sound /v/ is an exception: while identification of /v/ deteriorates in noise for all groups, there is no improvement in quiet for the consonant-trained groups and even a slight reduction in quiet for the vowel-trained cohorts. This may be due to its inherent maskability and confusability with /ð/ in noise, its similarity to Spanish /b/, which is often realised as a frictionless continuant, and it being orthographically-merged with “b” in Spanish spelling.

The origin of the matched-benefit of CONS-N training is spread across several consonants, but those that show the largest gains relative to CONS-Q training are the nasals /n, m, ŋ/ and the plosive /p/. These categories are well-identified in quiet but were seen to be highly vulnerable to masking (fig. 3) prior to training. The effect of CONS-N training on the nasals is mainly to reduce their manner confusions (e.g., /n/ and /l/ with /d, /m/ with /b/), while place confusions are more resistant to training.

In support of these observations, figure 5 displays the percentage of transmitted information (Miller and Nicely, 1955) for manner, place and voicing for the two consonant-trained groups. Transmitted information provides an idea of the influence of specific phonetic features on consonant identification in noise, measured as the proportion of information for a given feature that is available to the listener (see Ch. 10 of Loizou, 2007, for an example). All three features show significant group by condition interactions [manner: $F(2, 84) = 6.44, p < .01, \eta^2 = 0.03$; place: $F = 8.7, p < .001, \eta^2 = 0.05$; voicing: $F = 10.5, p < .001, \eta^2 = 0.05$]. Cohort CONS-Q exceeded CONS-N for place and voicing in the quiet condition, while CONS-N showed a higher transmission of manner and voicing in the SSN condition [FLSDs: manner = 1.7, place = 1.8, voicing = 2.8]. No significant differences between the groups were evident in the BAB condition for any feature.

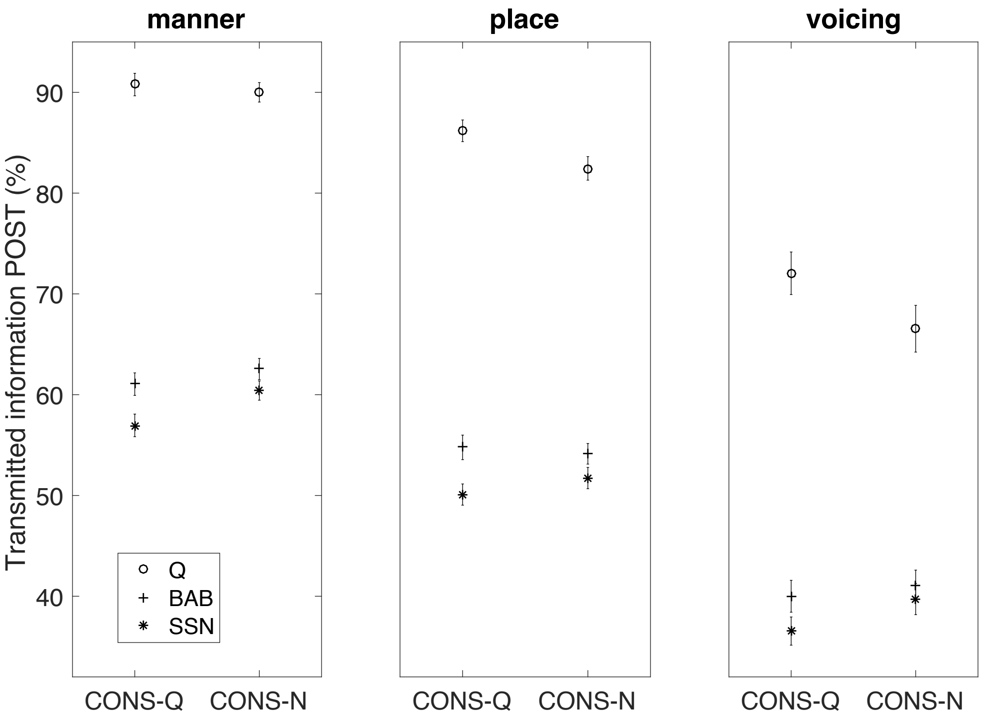

FIG. 5. Transmitted information for manner, place and voicing in the post-test for the consonant-trained groups.

D. Response times

Response times decreased for all groups and masking conditions between pre- and post-test, with post-test responses requiring between 70% and 86% of the time in the pre-test. However, no clear effect of differential training is evident in these results. A 3-factor ANOVA confirmed the lack of group effect [$p = 0.9$] and no two-way interactions of group with test phase nor masking condition (a marginally-significant 3-way interaction [$F(6, 164) = 2.28, p < .05; \eta^2 = 0.01$] can be ascribed to minor differences between the two consonant-trained groups on the BAB masker in the pre-test). The ANOVA confirms main effects of test phase [$F(1, 82) = 371; p < .001; \eta^2 = 0.40$] and masker condition [$F(2, 164) = 80.4; p <$

$.001; \eta^2 = 0.13$]. In the pre-test, listeners responded most rapidly to tokens presented in quiet and most slowly in SSN (quiet: 2664 ms; BAB: 2768 ms; SSN: 2911 ms; FLSD = 59 ms), with a similar ranking in the post-test (quiet: 1966 ms; BAB: 2297 ms; SSN: 2372 ms).

IV. DISCUSSION

Noise is present in many everyday speech communication scenarios, yet is a factor rarely considered in second language acquisition. The main goal of this study was to ascertain whether noise represents a barrier to non-native consonant acquisition. We considered the possibility that maskers might have a detrimental effect on acquisition due to the reduction in availability of cues to the identity of foreign language speech segments.

Four cohorts of Spanish learners underwent training regimes which differed in both the types of segments presented (vowels or consonants) and the presence or absence of masking noise, and their pre-to-post test improvements in English consonant identification were analysed. All listener groups showed improvements in the post-test. Gains for the groups trained on vowels provide a control measure of the perceptual benefits due to other factors such as vowel and consonant analysis and transcription practice which formed part of the module in English Phonetics that the participants were pursuing during the period of the experiment. Some incidental in-course learning effect was anticipated. Additionally, some of the identification gains may have been due to task habituation. In fact, the vowel-trained group gains from pre- to post-test are quite similar to the rapid gains observed between the pre-test and the first training session for the consonant-trained groups (fig. 2). The fact that such improvements occurred very early suggests that they were due to in-task accommodation, a form of procedural learning which is often observed in similar training paradigms (Robinson and Summerfield, 2006; Woods *et al.*, 2015), rather than resulting from exposure to the parallel course material, which would be expected to produce more gradual improvements.

In comparison to the modest improvements of around 2 to 4 percentage points exhibited by the vowel-trained groups, the two groups trained on consonants showed gains of between 10 and 14 percentage points. This outcome provides a clear demonstration that exposure to target consonants in noise during training is beneficial rather than harmful, relative to no exposure, since the cohort trained on consonants in noise showed significantly larger gains than either of the cohorts trained on vowel sounds. A comparison of the two consonant-trained groups also revealed a small but significant benefit worth around 2-3 percentage points when the training and test conditions matched: the cohort trained in quiet performed slightly better than the noise-trained group when tested in quiet, and conversely the group trained in speech-shaped noise showed larger gains when tested in that condition.

We found no evidence that habituation to specific details of the masker (cf. Wilson *et al.*, 2003) was responsible for some or all of the benefits of noise-based training. Exposure to masking noise during training on vowels did not lead to significantly larger gains for consonants presented in noise in comparison to a group trained on vowels in quiet conditions, suggesting that listeners were not merely learning to tune out the background or becoming familiar with the spectral properties of speech-shaped noise. However, on the basis of the current study we cannot entirely rule out the possibility of noise habituation since the level of masking noise required to have a significant impact on vowel identification is typically higher than that needed to reduce consonant categorisation accuracy, and although the vowel SNR was lower than the majority of the consonant SNRs during training, it is possible that listeners had no need to handle the masker in order to achieve good vowel recognition performance. Cognitive load measures (e.g., Gagné *et al.*, 2017; McGarrigle *et al.*, 2014) might reveal differences in the degree to which a given masking noise affects listeners even when intelligibility is near ceiling. While the current study did not measure cognitive load explicitly, we found no evidence of noise-training benefits in terms of faster response times, a measure which has been used as a proxy for listening effort (Pals *et al.*, 2015). A further limitation of the current study is the use of a single SNR during vowels-in-noise training. Although the SNR matched that of the consonant test SNR, the question of whether variation in the SNR might promote noise habituation merits further investigation.

We also hypothesised that exposure to a masker would benefit listeners by favouring the discovery of noise-robust cues, complemented by learning appropriate cue-weightings. This possibility is supported by the finding that the cohort trained on speech-shaped noise showed large gains when tested in 8-talker babble. However, gains in the babble condition were almost identical to those from the group trained on consonants in quiet. One interpretation of this outcome is that while both quiet and noise-based training are effective in handling a novel masker, the basis for the transfer is different in the two cases. In particular, masking leads to some loss of information, as demonstrated by the reduction in identification performance in noise, so those listeners who underwent noise-based training would have received incomplete spectro-temporal data as a consequence of masking, relative to those listeners who heard consonants in quiet conditions. However, the noise-trained group may have been able to compensate for the net loss of exposure by determining which information was reliable in the presence of a masker, something that those trained in quiet were unable to do. It is possible that the discovery of robust information compensated for the benefits of receiving intact spectro-temporal cues to consonants in the current study, but further work is required to investigate the mechanisms of transfer in the quiet and noise-trained cases.

We note that the highest levels attained by the consonant-trained groups are not far from native listener scores, which naturally represent a limit on performance. Indeed, gains asymptoted after around six training sessions, corresponding to around 120 exemplars per consonant. It is tempting to consider that further exposure would be irrelevant. However, longer training procedures have been seen as important for learning retention (e.g., Bradlow *et al.*, 1997; Woods *et al.*, 2015), something that we did not test in the current study.

V. CONCLUSIONS

Learning the sounds of a foreign language in the presence of noise is no barrier to their acquisition. Overall, listeners exposed to consonants in masking noise during an extensive training period exhibited improvements in identification rates similar to those for a group trained in quiet conditions. Both groups outperformed listeners trained on vowels in quiet or noise. A small matched-condition benefit was observed: noise exposure during training led to greater gains in noise than training in quiet, while conversely training in quiet produced larger gains in a noise-free test condition. We found no evidence that noise-habituation was responsible for these gains.

ACKNOWLEDGMENTS

This study was carried with funding from the Basque Government *Consolidados* grant to the Language and Speech Laboratory at the University of the Basque Country.

Bohn, O. S., and Flege, J. E. (**1990**). "Interlingual identification and the role of foreign language experience in L2 vowel perception," Applied Psycholinguistics **11**, 303–328.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tokhura, Y. (**1997**). "Japanese listeners to identify english /r/ and /l/: Iv. some effects of perceptual learning on speech production," Journal of the Acoustical Society of America **101**, 2299–2310.

Burk, M., Humes, L., Amos, N., and Strauser, L. (**2006**). "Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners," Ear and Hearing **27**, 263–278.

Cebrian, J. (**2006**). "Experience and the use of duration in the categorization of L2 vowels," Journal of Phonetics **34**, 372–387.

Clopper, C., and Pisoni, D. (**2004**). "Effects of talker variability on perceptual learning of dialects," Language and Speech **47**, 207–239.

Cooke, M., García Lecumberri, M. L., Scharenborg, O., and van Dommelen, W. A. (**2010**). "Language-independent processing in speech perception: identification of English intervocalic consonants by speakers of eight European languages," Speech Communication **52**, 954–967.

Cooke, M., and Scharenborg, O. (**2008**). “The Interspeech 2008 consonant challenge,” in *Proceedings of Interspeech*, pp. 1765–1768.

Cooper, A., Brouwer, S., and Bradlow, A. R. (**2015**). “Interdependent processing and encoding of speech and concurrent background noise,” Attention, Perception & Psychophysics **77**, 1342–1357.

Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (**2012**). “Word learning under adverse listening conditions: context-specific recognition,” Language and Cognitive Processes **27**, 1021–1038.

Cutler, A., García Lecumberri, M., and Cooke, M. (**2008**). “Consonant identification in noise by native and non-native listeners: effects of local context,” Journal of the Acoustical Society of America **124**, 1264–1268.

Florentine, M., Buus, S., Scharf, B., and Canevet, G. (**1984**). “Speech reception thresholds in noise for native and non-native listeners,” Journal of the Acoustical Society of America **75**, s84.

Gagné, J.-P., Besser, J., and Lemke, U. (**2017**). “Behavioral assessment of listening effort using a dual-task paradigm: a review,” Trends in Hearing **21**, 1–25.

García Lecumberri, M. L., and Cooke, M. (**2006**). “Effect of masker type on native and non-native consonant perception in noise,” Journal of the Acoustical Society of America **119**, 2445–2454.

García Lecumberri, M. L., Cooke, M., and Cutler, A. (**2010**). “Non-native speech perception in adverse conditions: A review,” Speech Communication **52**, 864–886.

Hayward, K. (**2002**). *Experimental Phonetics* (London: Pearson Education).

Humes, L. E., Burk, M. H., Strauser, L. E., and Kinney, D. L. (**2009**). “Development and efficacy of a frequent-word auditory training protocol for older adults with impaired hearing,” Ear and Hearing **30**, 613–627.

Kent, R. D., Dembowski, J., and Lass, N. (**1996**). “The acoustic characteristics of American English,” in *Principles of Experimental Phonetics*, edited by N. Lass (Mosby Yearbook), Chap. 5.

Kent, R. D., and Read, C. (**1992**). *The Acoustic Analysis of Speech* (San Diego: Singular Publishing Group).

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners,” Journal of the Acoustical Society of America **116**, 2395–2405.

Koziol, L. F., and Budding, D. E. (**2012**). “Procedural learning,” in *Encyclopedia of the Sciences of Learning*, edited by N. M. Seel and M. Norbert (Springer US, Boston, MA), pp. 2694–2696.

Lippmann, R., Martin, E., and Paul, D. (**1987**). “Multi-style training for robust isolated-word speech recognition,” in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 705–708.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). “Training Japanese listeners to identify English /r/ and /l/: A first report,” Journal of the Acoustical Society of America **89**, 874–886.

Loizou, P. (**2007**). *Speech Enhancement: theory and practice* (CRC Press).

Lovitt, A., and Allen, J. (**2006**). “50 years late: Repeating Miller-Nicely 1955,” in *Proceedings of Interspeech*, pp. 2154–2157.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). “Integration of multiple speech segmentation cues: a hierarchical framework,” J. Exp. Psych: General **134**, 477–500.

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., and Amitay, S. (**2014**). “Listening effort and fatigure: what exactly are we measuring,” International Journal of Audiology **53**, 433–445.

Miller, G., and Nicely, P. (**1955**). “Analysis of perceptual confusions among some English consonants,” Journal of the Acoustical Society of America **27**, 338–352.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,” Journal of the Acoustical Society of America **95**, 1085–1099.

Oba, S. I., Fu, Q.-J., and Galvin, J. J. (**2011**). “Digit training in noise can improve cochlear implant users’ speech understanding in noise,” Ear and Hearing **32**, 573–581.

Pals, C., Sarampalis, A., van Rijn, H., and Baskent, D. (**2015**). “Validation of a simple response-time measure of listening effort,” Journal of the Acoustical Society of America **138**, EL187–EL192.

Pufahl, A., and Samuel, A. G. (**2014**). “How lexical is the lexicon? Evidence for integrated auditory memory representations,” Cognitive Psychology **70**, 1–30.

Robinson, K., and Summerfield, A. Q. (**2006**). “Adult auditory learning and training,” Ear and Hearing **17**, 51–65.

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (**2012**). “Training to improve hearing speech in noise: biological mechanisms,” Cerebral Cortex **22**, 1180–1190.

Stecker, G. C., Bowman, G. A., Yund, E. W., Herron, J. J., Roup, C. M., and Woods, D. L. (**2006**). “Perceptual training improves syllable identification in new and experienced hearing-aid users,” Journal of Rehabilitation Research & Development **43**, 537–552.

Studebaker, G. (**1985**). “A rationalized arcsine transform,” Journal of Speech and Hearing Research **28**, 455–462.

Takata, Y., and Nabelek, A. (**1990**). “English consonant recognition in noise and in reverberation by Japanese and American listeners,” Journal of the Acoustical Society of America **88**, 663–666.

Van Dommelen, W. A., and Hazan, V. (**2010**). “Perception of English consonants in noise by native and Norwegian listeners,” Speech Communication **52**, 968–979.

Wilson, R. H., Bell, T. S., and Koslowski, J. A. (**2003**). “Learning effects associated with repeated word-recognition measures using sentence materials,” Journal of Rehabilitation Research & Development **40**, 329–336.

Woods, D. L., Doss, Z., Herron, T. J., Arbogast, T., Younus, M., Ettlinger, M., and Yund, E. W. (**2015**). “Speech perception in older hearing impaired listeners: benefits of perceptual training,” PLoS ONE **10**, e0113965.

Wright, R. (**2004**). “A review of perceptual cues and cue robustness,” in *Phonetically Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press), pp. 34–57.