Article:

Cooke, M., Garcia Lecumberri, M.L., Barker, J. orcid.org/0000-0002-1684-5660 et al. (1 more author) (2019) Lexical frequency effects in English and Spanish word misperceptions. Journal of the Acoustical Society of America, 145 (2). EL136-EL141. ISSN 0001-4966

https://doi.org/10.1121/1.5090196

Lexical frequency effects in English and Spanish word misperceptions

Martin Cooke

Ikerbasque (Basque Science Foundation), Bilbao, Spain

m.cooke@ikerbasque.org

María Luisa García Lecumberri

Language & Speech Lab, Universidad del País Vasco, Vitoria, Spain

garcia.lecumberri@ehu.es

Jon Barker

Department of Computer Science, University of Sheffield, Sheffield, UK

j.p.barker@sheffield.ac.uk

Ricard Marxer

Université de Toulon, Aix Marseille Univ, CNRS, LIS, Marseille, France

ricard.marxer@lis-lab.fr

(Dated: 15 January 2019)

Abstract: When listeners misperceive words in noise, do they report words that are more common? Lexical frequency differences between misperceived and target words in English and Spanish were examined for five masker types. Misperceptions had a higher lexical frequency in the presence of pure energetic maskers, but frequency effects were reduced or absent for informational maskers. The tendency to report more common words increased with the degree of energetic masking, suggesting that uncertainty about segment identity provides a role for lexical frequency. However, acoustic-phonetic information from an informational masker may additionally constrain lexical choice.

1. Introduction

It is a common experience for listeners to misperceive words under challenging conditions, but the manner in which degraded sensory evidence and prior language experience interact to produce the resulting 'slips of the ear' is poorly understood. One form of prior information that listeners might be forced to use in noise is lexical frequency. It has long been known that common words are more likely to be correctly recognised in noise than less frequent words[1,2] but there are conflicting findings as to whether *misperceptions* are themselves more common words than corresponding intended 'target' words in noise.

Several studies have examined lexical frequency effects in naturalistic compilations of reported real-life misperceptions[3–5]. Using a meta-corpus composed of previous compilations of misperceptions, Tang[6] found an inconsistent pattern of lexical frequency effects across corpora, but overall, misperceptions were not more common words than target words. One issue with naturalistic corpora is the paucity or absence of metadata describing the context in which each misperception occurred. For example, neither audio evidence for each speech token nor information about the presence, nature and level of any maskers is available for further analysis.

Very few studies have measured lexical frequency effects in controlled masking conditions. Pollack et al.[7] analysed incorrect responses from an earlier study[8] in which listeners identified 144 distinct monosyllabic English words belonging to one of eight frequency classes, presented in white noise at signal-to-noise ratios (SNRs) in the range -5 to $+25$ dB. Pollack et al. found the median lexical frequency of incorrect responses to be independent of the word frequency class of the stimulus. However, listeners reported higher frequency misperceptions at lower SNRs. Listeners in a study by Felty et al.[9] identified subsets of a 1428 English word sample presented in 6-talker babble at SNRs of 0, +5 and +10 dB. A clear lexical frequency difference effect was observed: misperceptions were more common words than target words. A similar study[10] using Spanish words presented in five maskers at SNRs in the range −13 to +1 dB found that, across maskers, the lexical frequency of misperceived words was significantly higher than target words. However, no breakdown by masker type was presented.

Taken together, previous studies present an inconsistent picture of whether misperceptions reported by listeners under conditions of actual or potential masking (the latter corresponding to the case of naturalistic corpora) tend to be more common than target words. This is not altogether unexpected, since the varied masking conditions employed in the these studies might have modulated the role of lexical frequency in different ways. One key distinction is between energetic and informational masking. While pure energetic maskers such as stationary or modulated noise act to distort or partially remove acoustic evidence for the target word, speech-based informational maskers can in principle contribute fragments of their own acoustic-phonetic information to the melange which forms the basis for a listener's lexical decision, enabling misperceptions to result from misallocation of masker fragments to the final word interpretation. The current study of lexical frequency effects in noise was motivated by the absence of prior studies involving comparisons of (i) more than one masker; and (ii) maskers with both an energetic and informational component.

2. Datasets

Lexical frequency effects were investigated for consistent word misperceptions in noise in two recent, extensive, open-source datasets of Spanish (SP)[11] and English (EN) words[12]. Consistent misperceptions are defined in these datasets as tokens for which no fewer than six listeners reported the same misperception in response to a given target word presented in noise. Both datasets were elicited in a similar manner, but while the EN dataset used three maskers (speech-shaped noise, SSN; 3-talker babble-modulated noise, BMN3; 4-talker babble, BAB4), the SP dataset additionally employed 1-talker babble-modulated noise (BMN1) and 8-talker babble (BAB8). Babble maskers were generated by random concatenation of target words to reach the required babble density[11,12]. Misperceptions were elicited at a range of SNRs (Table 1), values chosen in pilot tests to maximise the chance of consistent confusions, motivated by the finding that too-high SNRs lead to few errors, while too-low SNRs tend to produce inconsistent errors.

The online Spanish and English corpora contain 3235 and 3207 misperceptions respectively. For the current study these counts were reduced to 3126 and 3198 after excluding tokens based on the following criteria: (i) 82 Spanish confusions were found to result from SNRs outside the desired SNR range[11]; (ii) for 11 examples (8 Spanish) no lexical frequency data was available for the confused word; and (iii) for 25 examples (19 Spanish) the reported misperception was not present in the relevant pronunciation dictionary. Table 1 provides a breakdown of the number of misperceptions for each language/masker pairing along with details of the SNRs that led to the misperceptions.

Table 1. *Misperception counts for the Spanish (SP) and English (EN) datasets in each masking condition, alongside statistics of the SNRs used during their elicitation, which varied within the range shown. 'Unique' refers to counts after removing duplicates (see section 3.1).*

		Counts		SNRs (dB)			
Masker	Dataset	Total	Unique	mean	std.	min	max
SSN	SP	609	437	-5.4	0.9	-7.0	-4.0
	EN	1068	759	-5.5	0.9	-7.0	-4.0
BMN3	SP	732	533	-5.3	1.4	-7.9	-3.0
	EN	1196	903	-5.5	1.4	-8.0	-3.0
BMN1	SP	777	611	-9.9	1.8	-13.0	-7.0
BAB8	SP	419	345	-1.2	1.3	-4.0	1.0
BAB4	SP	589	501	-0.9	1.2	-3.0	1.0
	EN	934	818	-1.2	1.2	-3.0	1.0

One difference between the two published corpora lies in the source of lexical frequency estimates for target and misperceived words. Estimates for the SP dataset are derived from the CREA Spanish word frequency list[13], expressed in occurrences per million words, while equivalent data for the EN dataset come from the SUBTLEX-UK corpus[14], expressed in

Zipfs. The Zipf scale is defined as $log10$ (frequency per billion words) and ranges from around 1 (very low frequency words) to 7 (extremely common words, mainly function words and pronouns). For example, in the current datasets, common words “por” and “we” have a Zipf value around 7 and the far less common words “bromeas” and “fifteenth” have values near to 2. The Zipf scale is argued to avoid the problem of interpreting negative values that arise from log-transformed counts per million words that occur when counts are derived from very large corpora[14]. To ease comparability in the current study, lexical frequencies in the SP dataset were derived by converting values to Zipfs. Mean word frequencies for the SP and EN datasets are very similar, at 4.23 (std. 0.70) and 4.20 (std. 0.81) Zipfs respectively.

3. Results

3.1 Lexical frequency differences

Lexical frequency differences were computed by subtracting the frequency of the target word from that of the misperception, so that positive lexical frequency differences correspond to misperceptions that are more common words.

Across masking conditions, mean lexical frequency differences for SP and EN are 0.39 and 0.44 Zipfs respectively, indicating that, on average, misperceptions are 2.5-2.75 times more common than target words. A breakdown by type of masker (Fig. 1) reveals that the lexical frequency difference originates largely in the pure energetic maskers (SSN, BMN1, BMN3); the two maskers with an informational component (BAB4, BAB8) show a much smaller lexical frequency effect. For example, on average, Spanish misperceptions reported

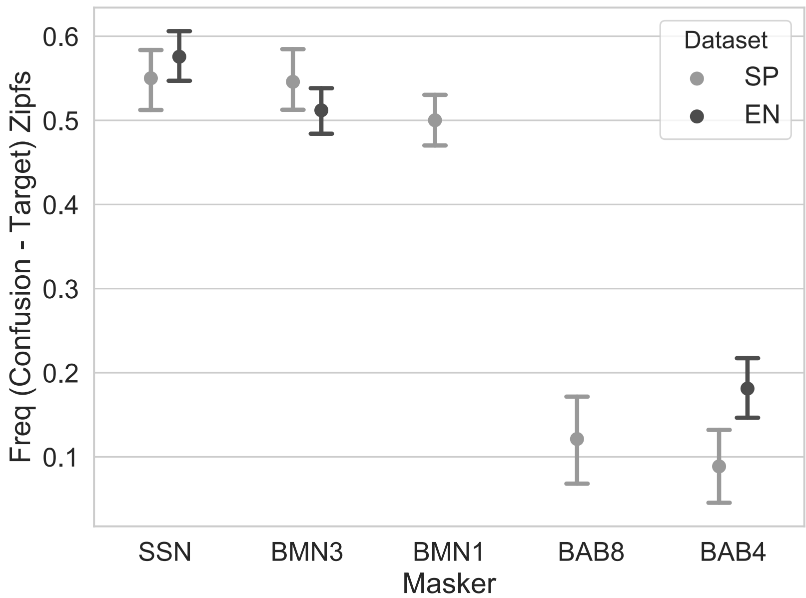

Fig. 1. Mean lexical frequency differences for each masker and dataset. Error bars indicate ±1 standard error. The Zipf scale is defined as $log10$ frequency per billion words (see text for details).

in SSN are over 3.5 times more common than their corresponding targets, while Spanish misperceptions reported in BAB4 occur only 1.2 times as frequently.

Lexical frequency differences for the three maskers in common in the SP and EN datasets are very similar. A two-factor analysis of variance based on the common maskers confirms a differential effect of masker type $[F(2, 5122) = 86.3, p < .001, \eta^2 = 0.033]$, a lack of effect of dataset $[p = .33]$ and the absence of an interaction between the two $[p = .20]$.

To test for any potential influence from the different sources of word frequency statistics used to compile the EN and SP datasets, lexical frequency counts for the SP dataset were replaced by those from the SUBTLEX-ESP corpus[15], which used a similar subtitle-based set of materials as those underlying the EN corpus[14]. A near-identical pattern [$F(2, 5122) = 89.3, p < .001, \eta^2 = 0.033$] was observed.

To eliminate further possible confounds, we examined four factors that might have influenced the pattern of lexical frequency differences across masker types.

3.1.1 Words from the masker

Misperceptions in the EN and SP datasets occasionally correspond to complete words contained in the BAB4 masker. Since maskers were constructed from the same set of speech materials as the target words, a mean lexical frequency difference of zero is to be expected for these cases, leading to a potential source of bias. Such cases amount to some 316 tokens (around 5% of the combined datasets), of which 269 occur in the EN dataset. After excluding these cases, the lexical frequency difference for the BAB4 masker in EN increases from 0.18 to 0.31 Zipfs, while for SP the increase is more modest, from 0.09 to 0.12 Zipfs. However, a significant masker effect remains [$F(2, 4806) = 47.1, p < .001, \eta^2 = 0.019$], indicating that the occasional reporting of complete words from the babble might account for part of the limited lexical frequency effect in the EN dataset, but has almost no impact on the SP corpus.

3.1.2 Word length differences

Shorter words tend to be more common, and different maskers may result in different patterns of phoneme deletion. For example, the quasi-stationary SSN masker might be expected to leave more energetic target components near to syllable nuclei intact, while maskers with significant temporal modulation might produce a more uniform pattern of deletions across phonemes.

For the misperceptions of the current datasets, length in phonemes is indeed inversely-related to lexical frequency [EN: $r = -0.38$, SP: $r = -0.30$, both $p < .001$]. However, a significant masker effect remains, albeit with a reduced effect size, after excluding target/misperception pairs of differing phoneme length [$F(2, 1999) = 21.2, p < .001, \eta^2 = 0.021$]. Combining the equal-length criterion with exclusion of words from the masker (Section 3.1.1) leads to a further reduction in effect size [$F(2, 1929) = 13.9, p < .001, \eta^2 = 0.014$]. Interpretation of causality in the relationship between lexical frequency and word length is problematic[9], since misperceptions may be shorter *because* they are of higher frequency.

3.1.3 Influence of extreme Zipf values

To check whether lexical frequency differences were influenced by extreme Zipf values, we examined the ratio of the number of target-misperception pairs with a positive lexical frequency difference to those with a negative lexical frequency difference, a metric that removes the influence of absolute frequency values. Across maskers, the ratio produces a clear bimodal pattern similar to that seen in Fig. 1. For example, about 2.8 times as many pairs have a positive difference for the SSN masker, a ratio that decreases to 1.3 for the BAB4 masker.

3.1.4 Duplicate target-misperception pairs

Due to the procedure used to generate new speech-in-noise tokens on demand during elicitation of the EN and SP datasets, which involved random selection of a target word from a base corpus, both datasets contain a number of duplicate target-misperception pairs (note that even though targets were presented multiple times, they may have come from different talkers and were mixed at varying SNRs with potentially different maskers). For instance, the target word ‘perverse’ (3.13 Zipfs) was misperceived as the more common word ‘reverse’ (4.15 Zipfs) on two occasions. A re-analysis limited to unique pairs only (counts of which are indicated in Table 1) produces a highly-similar pattern to that seen in Fig. 1 and a clear effect of masker type $[F(2,3945) = 70.0, p < .001, \eta^2 = 0.034]$, ruling out any influence from duplicate pairs.

3.2 Energetic masking

Although the lexical frequency differences observed in Fig. 1 vary across maskers, these differences might not stem from masker type *per se* but rather from differences in degree of energetic masking, which were not fixed or equalised across maskers (recall that SNR ranges were chosen to favour the elicitation of misperceptions for that masker type). As noted in the Introduction, one study[7] found that listeners were more likely to report words with a higher lexical frequency in noise at low SNRs. Our results also show a negative correlation between the size of the lexical frequency difference and SNR across maskers [SP: $r = -0.15$, EN: $r = -0.12$; both $p < .001$]. However, SNR is known to be a poor predictor of intelligibility when comparing maskers which vary in their spectro-temporal modulation properties[16]. As an alternative proxy for pure energetic masking, glimpse percentages were computed for the two datasets (Fig. 2). Glimpse percentages have been shown to provide reasonable first-order predictions of intelligibility for a range of different speech and masker types[17].

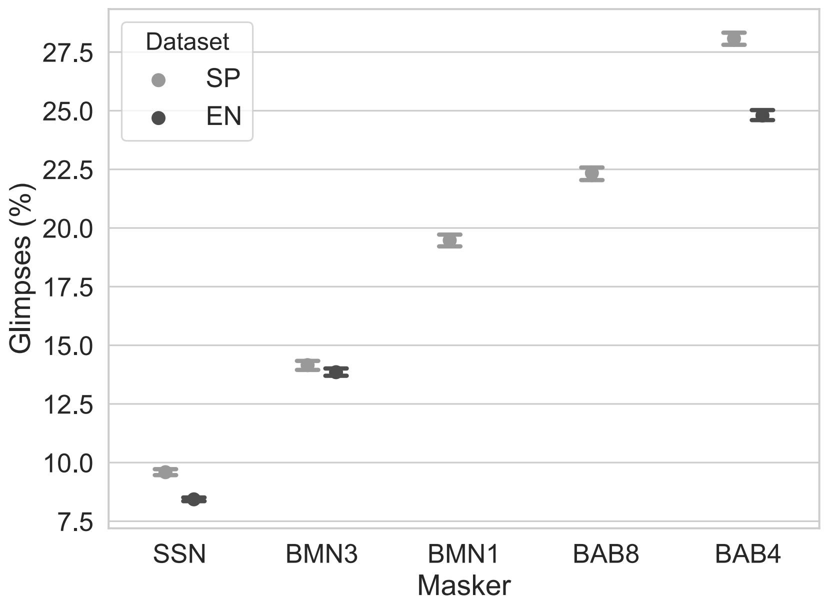

Fig. 2. Mean glimpse percentages for each dataset and masker combination. Glimpse percentage is defined here as the percentage of spectro-temporal regions in an auditory ‘spectrogram’ where the target word energy exceeds that of the masker. Auditory spectrograms were computed by processing the target word and masker independently through a 55-channel gammatone filterbank with centre frequencies in the range 80-8000 Hz, followed by extraction of the Hilbert envelope, smoothing with a 0.8 ms time constant, and downsampling to 100 Hz. Error bars indicate ±1 standard error.

Mean glimpse percentage differs across maskers $[F(2, 5122) = 4711, p < .001, \eta^2 = 0.64]$ for the ranges of SNRs used here. Moreover, glimpse percentage is lower for the three pure energetic maskers than for the two babble maskers, suggesting that part of the lexical frequency effect seen in Fig. 1 could be due to scarcity of information about the target that is predicted to survive masking. Nevertheless, energetic masking cannot entirely explain the across-masker disparity between lexical frequency differences. For example, BMN1 and BAB8 are both predicted to leave the target word occupying around 20-22% of the spectro-temporal plane, yet the BMN1 masker results in a far larger lexical frequency effect.

4. Discussion

In two extensive corpora, misperceptions reported by listeners were words of a higher lexical frequency than their corresponding intended target words. Lexical frequency difference shows an apparent dependence on the type of masker, being substantially larger for three pure energetic maskers than two babble maskers (Fig. 1), but some of the effect may be due to differences in the amount of acoustic information which survives masking according to a glimpsing model (Fig. 2). This outcome supports the finding of an increased lexical frequency effect at lower SNRs[7]. It is conceivable that increased acoustic uncertainty favours the use of word frequency priors. An example from the EN dataset illustrates this possibility: "clinic" with a lexical frequency of 3.9 Zipfs was misperceived as "finish" (5.1 Zipfs) in the presence of the SSN masker, perhaps due to the masker eroding acoustic-phonetic information for the target word apart from evidence for the two vowel nuclei, leaving listeners to hypothesise a word with the corresponding vowels. In such situations, one would anticipate lexical frequency having a role in the choice of word to report. If lexical frequency is more likely to come into play in more adverse masking conditions, the absence of a frequency effect for naturalistic word misperceptions[6] may be due to the environment under which misperceptions occurred being insufficiently adverse to engage prior lexical frequency information in the process of deciding which word was heard.

However, a comparison of Figs. 1 and 2 suggests that something more than pure energetic masking is needed to fully explain the role of lexical frequency in word misperceptions. One possibility is that the additional acoustic-phonetic information contributed by an informational masker acts as a source of constraint on possible word hypotheses, which in turn limits or eliminates a role for lexical frequency. This notion can be illustrated with an example from the EN dataset: target word "wife" (5.2 Zipfs) was misperceived in BAB4 for the less frequent word "twice" (4.8 Zipfs); an inspection of the words making up the babble provides clear evidence for a word-initial /t/ and a word-final /s/ with a temporal alignment appropriate for their incorporation in the reported word. Here, the ability to fit alternative word candidates is constrained by elements in the babble, attenuating the influence of lexical frequency. The limited room for manouevre in the presence of additional phonetic cues contrasts with the uncertainty created in the face of missing information due to pure energetic masking. Further support for this hypothesis awaits a detailed examination of each individual misperception in the context of the acoustic information of the babble signal which elicited the misperception.

The reduced lexical frequency effect for our 4- and 8-talker babble maskers for SNRs below 0 dB is at odds with the findings of Felty et al.[9], who reported a lexical frequency effect for a 6-talker babble masker for SNRs in the range 0-10 dB. However, there are differences in both the speech and masker materials and the elicitation techniques used in the two studies.

Perhaps the biggest disparity is in the mean lexical frequency of the target words. Using the US subtitle-based lexical frequency data[18], we calculated the mean Zipf value for the target words of Felty et al. to be 3.16 Zipfs, a value substantially lower than the mean of around 4.2 Zipfs for the datasets of the current study. Since low frequency targets are *a priori* more likely to result in higher frequency responses than targets of a higher mean lexical frequency, it is understandable that Felty et al.[9] observed a lexical frequency effect at higher SNRs than those used in the current study. A further difference between the two studies is the nature of the speech material making up the babble. In Felty et al.[9] the masking material came from a different talker than that of the target words, while in the current study the target talker could also appear in the babble. Informational masking effects are thus expected to be higher for our stimuli, and as a consequence it seems likely that speech fragments from the masker were more easily misallocated into the final word misperception.

The structure of English and Spanish differs in many respects, including vowel inventory size (greater for EN), inflectional morphology (richer for SP), and presence of consonant clusters (greater for EN). In spite of these differences, the similar across-masker patterning of lexical frequency effects for the two languages suggests that relatively low-level processes such as energetic masking and misallocation of acoustic-phonetic evidence from the masker can modulate the extent to which lexical frequency priors are engaged during word recognition.

5. Conclusions

Across five types of masker and two languages, listeners reported words of a higher lexical frequency than the intended target words. The size of the lexical frequency effect was larger for pure energetic maskers than for maskers containing speech. However, the pure energetic maskers of the current study possessed a greater predicted masking potential than the babble maskers, suggesting that lexical frequency has more influence when acoustic-phonetic information is scarce. The role of lexical frequency might be reduced in the presence of a speech-based masker, by limiting the number of lexical hypotheses compatible with audible acoustic-phonetic evidence from both target and masker.

Acknowledgments

This work received support from the Basque Government under grant *Language and Speech* (IT311-10).

References and links

[1]D. Howes, "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**(2), 296–305 (1957).

[2]H. B. Savin, "Word-frequency effect and errors in the perception of speech," J. Acoust. Soc. Am. **35**(2), 200–206 (1963).

[3]C. Browman, "Perceptual processing: Evidence from slips of the ear," in *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, edited by V. A. Fromkin (New York: Academic Press, New York, 1980).

[4]Z. Bond, "Morphological errors in casual conversation," Brain and Language **68**(12), 144–150 (1999).

[5]K. Tang and A. Nevins, "Measuring segmental and lexical trends in a corpus of naturalistic speech," in *Proceedings of the 43rd Meeting of the North East Linguistic Society*, edited by H.-L. Huang, E. Poole, and A. Rysling, Vol. 2 (GLSA (Graduate Linguistics Student Association), 2014), pp. 153–166.

[6]K. Tang, "Naturalistic speech misperceptions," Ph.D. thesis, UCL, 2015.

[7]I. Pollack, H. Rubenstein, and L. Decker, "Analysis of incorrect responses to an unknown message set," J. Acoust. Soc. Am. **32**, 454–457 (1960).

[8]I. Pollack, H. Rubenstein, and L. Decker, "Intelligibility of known and unknown message sets," J. Acoust. Soc. Am. **31**(3), 273–279 (1959).

[9]R. Felty, A. Buchwald, T. Gruenenfelder, and D. Pisoni, "Misperceptions of spoken words: Data from a random sample of American English words," J. Acoust. Soc. Am. **134**, 572–585 (2013).

[10]M. Tóth, "A microscopic analysis of consistent word misperceptions," Ph.D. thesis, University of the Basque Country, 2017.

[11]M. A. Tóth, M. L. García Lecumberri, Y. Tang, and M. Cooke, "A corpus of noise-induced word misperceptions for Spanish," J. Acoust. Soc. Am. **137**, EL184–EL189 (2015).

[12]R. Marxer, J. Barker, M. Cooke, and M. L. García Lecumberri, "A corpus of noise-induced word misperceptions for English," J. Acoust. Soc. Am. **140**, EL458–EL463 (2016).

[13]Real Academic Española, "Corpus de Referencia del Español Actual (CREA)," http://www.rae.es/recursos/banco-de-datos/crea (2008).