



HAL
open science

Consonant bias in adult lexical processing under acoustically degraded listening conditions

Irene de la Cruz-Pavía, Coraline Eloy, Paula Perrineau-Hecklé, Thierry Nazzi, Laurianne Cabrera

► **To cite this version:**

Irene de la Cruz-Pavía, Coraline Eloy, Paula Perrineau-Hecklé, Thierry Nazzi, Laurianne Cabrera. Consonant bias in adult lexical processing under acoustically degraded listening conditions. *JASA Express Letters*, 2023, 3 (5), pp.055206. 10.1121/10.0019576 . hal-04197011

HAL Id: hal-04197011

<https://hal.science/hal-04197011v1>

Submitted on 5 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Consonant bias in adult lexical processing under acoustically degraded listening conditions

Irene de la Cruz-Pavía,^{1,a)}  Coraline Eloy,² Paula Perrineau-Hecklé,² Thierry Nazzi,² 
and Laurianne Cabrera² 

¹Department of Linguistics and Basque Studies, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Vitoria-Gasteiz 01006, Spain

²Integrative Neuroscience and Cognition Center, Université Paris Cité, Centre National de la Recherche Scientifique, Paris 75006, France

idelacruzpavia@gmail.com, coraline.elay@gmail.com, paulaperrineau@gmail.com, thierry.nazzi@parisdescartes.fr, laurianne.cabrera@parisdescartes.fr

Abstract: Consonants facilitate lexical processing across many languages, including French. This study investigates whether acoustic degradation affects this phonological bias in an auditory lexical decision task. French words were processed using an eight-band vocoder, degrading their frequency modulations (FM) while preserving original amplitude modulations (AM). Adult French natives were presented with these French words, preceded by similarly processed pseudoword primes sharing their vowels, consonants, or neither. Results reveal a consonant bias in the listeners' accuracy and response times, despite the reduced spectral and FM information. These degraded conditions resemble current cochlear-implant processors, and attest to the robustness of this phonological bias. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Douglas D O'Shaughnessy]

<https://doi.org/10.1121/10.0019576>

Received: 6 March 2023 Accepted: 5 May 2023 Published Online: 23 May 2023

1. Introduction

Current auditory rehabilitation technologies capitalize on psychoacoustic models that describe how the typical auditory system extracts the acoustic components of speech. This processing is thought to result from the action of a series of filters extracting spectral information from base to apex within the cochlea. At the output of each cochlear filter, sounds are then modelled as a series of narrowband signals modulated in amplitude over time. Two time scales have been identified in auditory processing: a relatively slow one, called amplitude modulation (AM or temporal envelope), and a faster one corresponding to the variations in instantaneous frequency close to the center of the frequency band, called frequency modulation [FM or temporal fine structure; see Moore (2004)].

Cochlear implants (CIs) are offered to individuals with profound sensorineural hearing loss to restore hearing to a certain level via electrodes implanted within the cochlea. CI processors deliver the AM cues of the original signal via a limited number of (relatively) broad spectral channels (Shannon, 2012). Acoustic simulation results compared to speech perception performance of CI users indicate that the independent number of channels for information transmission varies between four and eight channels (Friesen *et al.*, 2001; Fu and Nogaki, 2005). Moreover, CI processors do not explicitly transmit the FM cues of the original signal; these are instead replaced by a fixed train of electrical pulses whose amplitude is modulated by the original AM cues. As a result, the output signal delivered by CIs is severely impoverished in both spectral and fine temporal resolution compared to the original signal.

A wealth of studies aiming at simulating CI processing in listeners with normal hearing have used speech analysis algorithms, called vocoders, to selectively extract the temporal modulations of original speech (AM + FM) in a given number of spectral bands. By reducing the number of spectral bands and/or selectively degrading the AM/FM components in each band, one can assess the specific role of spectral and temporal information in speech perception for normally hearing participants (Drullman *et al.*, 1994; Shannon *et al.*, 1995; Smith *et al.*, 2002). This body of work shows that normal-hearing listeners obtain near perfect word identification scores in quiet when FM cues are reduced and only AM cues are preserved in 4 broad spectral bands. Thus, listeners with normal hearing are able to rely only on slow temporal cues to correctly identify speech in quiet, and succeed under conditions that simulate current CI processors (Shannon *et al.*, 1995).

Interestingly, the effect of vocoder degradation (spectral cues, FM and AM cues) appears to affect differently the performance of vowel and consonant identification. Consonant identification appears to be less affected by spectral degradation than vowel identification, with relatively good performance observed when only 8 bands were preserved for

^{a)} Author to whom correspondence should be addressed. Also at: Basque Foundation for Science Ikerbasque, Bilbao 48009, Spain.

consonants, in contrast to the 12 bands necessary to achieve similar performance with vowels (Xu *et al.*, 2005). The opposite pattern is observed for temporal degradation, as filtering out AM cues above 16 Hz drastically affects consonant identification, while a comparable impact is first observed at a cut-off of 4 Hz for vowels (Xu *et al.*, 2005). These vocoder studies with normal-hearing listeners usually consist of sentence recognition tasks (Shannon *et al.*, 1995) or phoneme recognition tasks that allow extracting confusion matrices of target phonemes (Xu *et al.*, 2005; Xu and Zheng, 2007). Importantly, these previous psychoacoustic studies have not yet examined the role of a phonological bias well attested in psycholinguistic research: the facilitatory role of consonants as compared with vowels in *lexical processing*.

Consonants and vowels have been proposed to play distinct roles in language processing, with consonants being more important than vowels in lexical processing and vowels being more important than consonants in prosodic-syntactic processing (Nespor *et al.*, 2003). The phonological bias known as *consonant bias in lexical processing* has been observed in numerous languages (though possibly not all, see data on Mandarin and Cantonese) and across auditory, or phonologically related tasks [see Nazzi and Cutler (2019) for a review]. Indeed, when learning new pairs of pseudo-words, adult listeners confuse two pseudo-words more often when they share the same consonants than when they share the same vowels (Creel *et al.*, 2006; Escudero *et al.*, 2016; Havy *et al.*, 2014). When asked to change the phonemes of a pseudo-word to obtain a real word, listeners are faster and change more often its vowels compared to its consonants (Van Ooijen, 1996). In segmentation tasks of continuous speech using artificial languages, listeners are better at extracting words when transitional probabilities link common consonants than common vowels (Bonatti *et al.*, 2005). In lexical decision tasks, in which listeners are asked to determine whether a word presented auditorily is a real word or not, the presence of an auditory pseudo-word prime sharing consonants with the target word decreases response times compared to vowels (Delle Luche *et al.*, 2014). This facilitatory effect of consonants for lexical processing emerges very early in development [in the first 1 or 2 years of life, although the exact timing varies across languages (Nazzi and Cutler, 2019)], and has been proposed to bootstrap lexical acquisition (Nishibayashi and Nazzi, 2016).

To date, the robustness of this consonant bias in lexical processing has not been tested under acoustically degraded conditions for listeners with normal hearing. The current study thus examines whether this bias survives specific acoustic degradations of fine spectro-temporal modulations that imitate the listening conditions of individuals with sensorineural hearing loss using CIs. Note that while AM-vocoders with reduced spectral resolution simulate CI processors (Friesen *et al.*, 2001; Zeng *et al.*, 2005), performance of normally hearing listeners cannot be entirely generalized to hearing impaired listeners using CIs, as further auditory processing may be affected along the auditory pathway in the case of sensorineural hearing loss. Nonetheless, it is essential to determine whether this phonological bias is observed under acoustically degraded speech conditions, given its well-established role in lexical acquisition and processing in normal-hearing infants and adults.

In the present study, we test whether fine spectro-temporal degradation affects the consonant bias in a lexical decision task using auditory priming. To do this, we replicated a lexical decision study by Delle Luche *et al.* (2014), in which a consonant bias in French adult listeners is reported when presented with non-degraded, intact speech. They presented listeners with targets with two possible syllabic structures, CVCV or VCVC (where C: consonant, V: vowel), preceded by one of three possible types of primes: primes sharing only the consonants with their targets, primes sharing only their vowels, or totally unrelated primes sharing neither. Delle Luche and colleagues found faster response times in targets preceded by primes with shared consonants, i.e., a facilitatory effect of consonants or consonant bias, although, it was limited to targets with a VCVC structure. Meanwhile, a facilitatory effect of vowels (i.e., faster response times) caused by a rhyme bias was found in CVCV targets. In addition, they found faster response times, i.e., an overall facilitatory effect, in VCVC words.

We assessed the effect of *fine spectro-temporal* degradation on the robustness of the consonant bias in word recognition. Specifically, we degraded the original stimuli of Delle Luche *et al.* (2014) by extracting AM cues in eight broad spectral bands and replacing FM cues with a pure tone in each band [tone-excited vocoders have been shown to distort less AM cues than noise-excited vocoders (Kates, 2011)]. This manipulation is thought to roughly simulate current CI processors. If, as found in sentence repetition and syllable identification tasks (Shannon *et al.*, 1995; Xu *et al.*, 2005), such degradation of FM cues and spectral resolution affects identification of vowels to a greater extent than consonants, and given the consonant bias found in Delle Luche *et al.* (2014) with intact speech, we expected to observe a consonant bias in this experiment.

2. Methods

2.1 Participants

Forty-two native French speakers (24 females, 18 males; mean age: 23.55 years, range 20–30 years) took part in the experiment, which was conducted at the Université Paris Cité's Integrative Neuroscience and Cognition Centre. All participants reported no language or hearing impairment and received a 5 € compensation for their participation. Informed consent forms were obtained from all participants.

2.2 Stimuli and design

Stimuli consisted of the set originally created and used by Delle Luche *et al.* (2014), which comprised 192 disyllabic items. Of these, 48 were French nouns selected from the database LEXIQUE 3.70 (New *et al.*, 2001). Half had a CVCV structure (C: consonant, V: vowel; e.g., *dépôt*), the remaining half a VCVC structure (e.g., *endive*). These subsets were matched in their frequency, phonological and orthographic Levenshtein distances and unique points [see Delle Luche *et al.* (2014), Appendix 1]. Another 48 items consisted of distracter words also selected from LEXIQUE, which had the same proportion of tokens beginning with consonants and vowels but differing phonological structures (e.g., *facteur*, *objet*). The remaining 96 items consisted of non-words generated by the trigram tool of the LEXIQUE Toolbox (New *et al.*, 2001). These non-words had the same proportion of C- and V-tokens and respected French phonotactics (e.g., *banto*, *piskure*, *ebis*, *inpu*).

During the study, the 192 targets were preceded by primes (see Fig. 1). A third of the primes shared only their vowels with the target (i.e., vowel-related condition, henceforth V-related; e.g., prime: *zérai/zeʁe/*, target: *délai/delɛ/*). Another third shared only their consonants with the target (i.e., consonant-related condition, henceforth C-related; e.g., prime: *dipou /dipu/*, target: *dépôt /depo/*). The remaining third did not share any phonemes with their target (i.e., unrelated condition; e.g., prime: *datu /daty/*, target: *bandit /bãdi/*). None of the primes were real French words and the transformation of the phonemes corresponded to one (82.3%) or two feature changes (17.7%). Distribution of the three prime types was balanced across targets, distractors and non-words. All stimuli were recorded by a French native speaker (at a sampling rate of 22.050 kHz, with 16 bit resolution). In summary, the study comprised 6 conditions (3 prime types \times 2 syllabic structures), with 32 items per condition. However, following Delle Luche *et al.* (2014), we analyzed only the 48 target real French words (8 targets per condition \times 3 prime types \times 2 syllabic structures).

In order to alter the spectro-temporal modulations of the stimuli, we processed them with a vocoder similar to the one used in Cabrera *et al.* (2014), degrading their spectral resolution and FM cues. Spectral resolution was reduced by passing each speech sound through a bank of 8, fourth-order gammatone filters, each 4- ERB_N -wide with center frequency uniformly spaced along an ERB_N -number scale ranging from 80 to 8020 Hz. The ERB_N scale (Glasberg and Moore, 1990; Moore, 2003) allows simulation of the bandwidth of cochlear filters of the normal ear. ERB_N corresponds to the average equivalent rectangular bandwidth of the auditory filter, as determined using young normally hearing listeners tested at moderate sound levels (Moore, 2007). The goal of the present study was not to replicate the frequency map of CI devices, which may differ between brands and users, but rather to use an estimate of normal cochlear filters.

The Hilbert transform was applied in each bandpass filtered speech signal to extract the AM and FM components. The AM was lowpass-filtered using a zero-phase Butterworth filter [36 dB/octave roll-off, see Ardoint and Lorenzi (2010) and Cabrera *et al.* (2014)] and a cut-off frequency of $ERB_N/2$. The ERB_N corresponded to that of the normal cochlear filter tuned to the geometric center of the 4- ERB_N wide gammatone filter. The original FM carriers were replaced

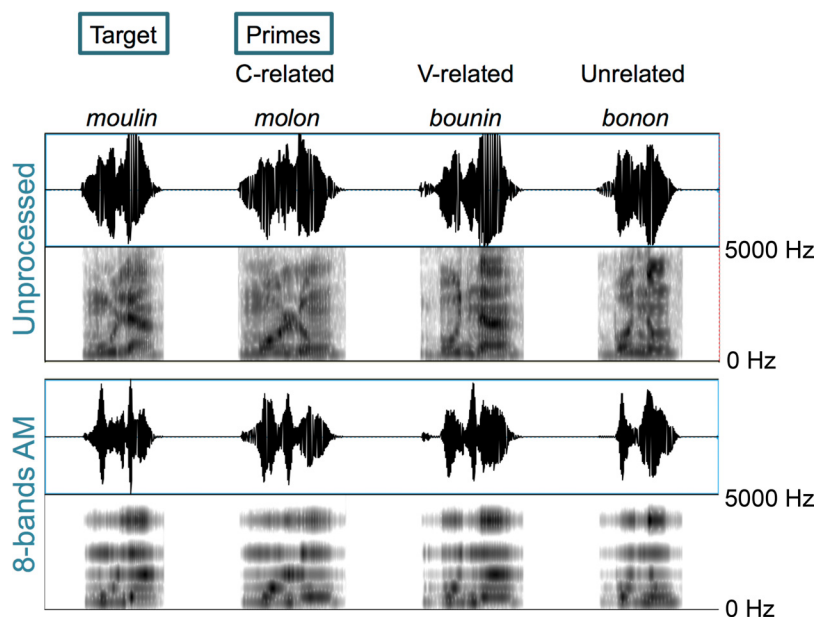


Fig. 1. Spectrograms depicting the acoustic degradation of the stimuli. The figure depicts one of the 48 targets used in the study (*moulin*, /muli/) and its 3 corresponding pseudo-word primes: a C-related prime, in which consonants are preserved but vowels changed (*molon*, /molõ/), a V-related prime, in which vowels are preserved but consonants changed (*bounin*, /buni/), and an unrelated prime in which both consonants and vowels are changed (*bonon*, /bonõ/). The upper panel depicts their corresponding spectrograms in unprocessed, intact speech, as used by Delle Luche *et al.* (2014), while the lower panel depicts their fine spectro-temporally degraded version used in the present experiment.

by sine wave carriers with frequencies at the center frequency of the gammatone filters, and with random starting phase in each analysis band. In each band, the new carrier was multiplied by the filtered AM function. The eight modulated speech signals were then added up and the level of the resulting speech signal was equalized in root mean square value as the input signal.

2.3 Procedure

Participants were tested individually in a quiet room and stimuli were presented using E-Prime 2.0 software (Psychology Software Tools) through headphones calibrated at 65 dB SPL. The study consisted of 192 trials and followed the procedure of [Delle Luche et al. \(2014\)](#). Thus, trials began with a fixation cross-presented at the center of the screen (500 ms). Participants then heard the prime, followed by the target (ISI: 10 ms). The trial ended 1500 ms after target offset or when the participant provided a response. Participants' task consisted of determining whether the second "word" they heard in each trial was or not a real French word, by pressing one of two buttons. The study began with 12 practice trials during which participants received feedback of their accuracy and response times. We used the three lists created by [Delle Luche et al. \(2014\)](#), in which prime-target pairs rotated (pseudo Latin-square design) to ensure that a given target was primed by only one prime condition in each list, but by all three conditions across the three lists. Items were pseudo-randomized within lists, so that no more than three words or non-words appeared consecutively. Each participant was presented with a single list. Therefore, each list was presented to 14 participants. Participants could take a pause after the first 96 trials, and the experiment had an average total duration of 20 min.

3. Results

3.1 Accuracy

Following [Delle Luche et al. \(2014\)](#), this and all subsequent analyses were limited to the 48 target real French words (for a total of 2016 data points: 48 items \times 42 participants). While at the group level all targets were tested in the 3 prime conditions, each participant contributed only 8 targets per condition (6 conditions: 3 prime types \times 2 syllabic structures), due to the study's pseudo Latin-square design (for a total of 336 data points per condition).¹ All analyses were conducted in R (version 4.1.1., [R Core Team, 2019](#)). We fitted generalized linear mixed effects models (lme4 package) with logit link. We began by fitting the conceptually most complex model, which explained the binomial dependent variable Accuracy (correct vs incorrect), with Prime type (consonant-related, vowel-related, unrelated), Syllabic Structure (CVCV vs VCVC) and their interaction as fixed effects, Subject and Item as random factors, and allowed Prime type and Syllabic Structure to vary randomly by Subject, and Prime type to vary by Item. We then built decreasingly complex models, running analyses of variances (ANOVAs) to compare pairs of models. The simplest model that fitted the data (i.e., the final model²) included the fixed effects of Prime type and Syllabic Structure, and the random factors of Subject and Item. The goodness of fit of the model was confirmed using the diagnostics available in the DHARMA library ([Hartig, 2022](#)). Using the function Anova (car package), we observed significant effects of Prime type [$\chi^2(2) = 7.66, p = 0.022$] and Syllabic Structure [$\chi^2(1) = 5.08, p = 0.024$], which we further analyzed using marginal means (emmeans package) and Holm correction for multiple comparisons whenever applicable.

Analysis of Prime type showed that participants' accuracy was significantly above chance in all three conditions ($p \leq 0.001$ for all three conditions), but revealed significantly greater accuracy in targets with C-related primes (69%) as compared with unrelated primes (63.7%) (OR = 0.708, SE = 0.092, z-ratio = -2.653, $p = 0.024$). Comparison of targets with C- and V-related (64.4%) primes was not significant (OR = 0.766, SE = 0.100, z-ratio = -2.052, $p = 0.080$), and neither was the comparison of targets with V-related and unrelated primes (OR = 1.082, SE = 0.138, z-ratio = 0.617, $p = 0.537$). In turn, analysis of Syllabic Structure showed that accuracy was significantly above chance in both structures (both $p \leq 0.001$), but revealed significantly greater accuracy in VCVC targets (71.3%) as compared with CVCV targets (60.1%) (OR = 0.559, SE = 0.150, z-ratio = -2.173, $p = 0.030$) (see Fig. 2).

3.2 Response times (RTs)

As in [Delle Luche et al. \(2014\)](#), incorrect responses ($k = 691, 34.28\%$) and outliers (i.e., RTs greater than ± 2.5 SD from the grand and individual means, $k = 25, 1.24\%$) were discarded from analysis. The remaining 1300 data points (64.48%) entered the analysis. We fitted linear mixed effects models (lme4 package). We first fitted the conceptually most complex model (identical to the one reported in the accuracy analysis above), and then step-by-step fitted decreasingly complex models. The model that best fitted the data (i.e., the final model³) included the interaction between Prime type and Syllabic Structure, and the random factors Subject and Item (RMSE = 163 ms). The function Anova revealed significant effects of Prime type [$\chi^2(2) = 11.80, p = 0.003$] and Syllabic Structure [$\chi^2(1) = 12.43, p < 0.001$] and their interaction [$\chi^2(2) = 7.89, p = 0.019$]. We analyzed this interaction further using marginal means (emmeans package, R) and Holm correction for multiple comparisons, which revealed slower response times to targets with V-related primes (mean: 1081 ms), that is, primes with different consonants than their targets, as compared with C-related [mean: 1027 ms, $t(1226) = 3.970, p < 0.001$] and unrelated primes [mean: 1053 ms, $t(1220) = 2.800, p = 0.026$], but only in targets with a VCVC structure. All other comparisons were not significant ($p \geq 0.319$) (see Fig. 2).

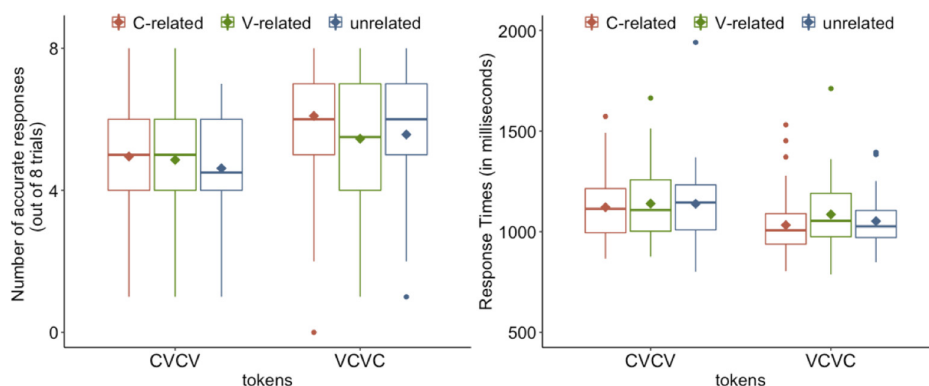


Fig. 2. Mean accuracy (left) and response times (right). Left panel: The box-and-whisker plots depict participants' median accuracy per condition (depicted by the central line within each box-plot), out of 8 trials, as well as the group's upper and lower quartiles. Right panel: The Box-and-whisker plots depict participants' median response times in trials with correct responses and the corresponding quartiles. Within plots, accuracy/response times are presented for each of the 3 Prime type conditions (C-related, V-related, unrelated, in red, green and blue box-plots, respectively) for both CVCV and VCVC tokens. The diamond shape within box-plots depicts the mean.

3.3 Summary of results

The results of the accuracy and response times analyses suggest, as predicted, the presence of a consonant bias in spite of the fine spectro-temporal degradation of the stimuli. Participants: (1) were more accurate at identifying whether a target was a real French word or not when it was preceded by a prime sharing the same consonants as compared with a prime sharing no segments, (2) responded more slowly when targets were preceded with primes with different consonants as compared with primes with different vowels and unrelated primes, although this was limited to targets with a VCVC structure. In addition, we found a facilitatory effect of VCVC targets, as revealed by participants' greater accuracy compared to CVCV tokens.

4. Discussion

In an auditory priming lexical decision study, we investigated whether a consonant bias, found in a lexical decision task using intact speech (Delle Luche *et al.*, 2014), is also observed when replicating the same experiment in acoustically degraded speech conditions. French native adults listened to speech with degraded spectral and temporal FM cues. Previous literature on other areas of speech perception reports that spectral degradation particularly impacts vowel identification, as compared with consonant identification (Xu *et al.*, 2005). Therefore, we predicted a consonant bias. The results fulfilled our predictions: although overall accuracy was lower than with the original stimuli [4.76% errors in Delle Luche *et al.* (2014) vs 34.26% in the present study], we found greater accuracy—i.e., a facilitatory effect—for targets *sharing* their consonants with the primes, and slower response times—i.e., a hindering effect—when targets and primes had *different* consonants.

The consonant bias is argued to play an important role in lexical processing and acquisition in normal-hearing individuals in numerous languages including French. Our results reveal—for the first time—that this bias is preserved for normally hearing listeners under acoustically degraded speech conditions that simulate cochlear-implant processors of individuals with sensorineural hearing loss. Hence, the acoustic information provided to CI users might be enough for post-lingually deafened adults to have the same consonant bias facilitating lexical processing as normally hearing adults, although direct evidence should be gathered on this issue in the future. If confirmed, it would be interesting to determine whether poor performers rely less on this phonological bias. Furthermore, investigating potential effects of electrode-neuron interfaces (e.g., electrode-neuron distance, intracochlear resistance, neural health) on this bias would also be of high interest, as variations in their quality have been shown to impact the identification of place of articulation as well as of vowels (Arjmandi *et al.*, 2022; DiNino *et al.*, 2016). Such individual evaluations may lead to a better identification of the phonetic features not well perceived by CI users, and their perceptual strategies for word recognition. These results may in turn inform current speech and language therapies, as well as signal processing strategies.

The results of the present experiment overall establish that the consonant bias found by Delle Luche *et al.* (2014) using intact speech extends to conditions of speech with degraded spectral and FM cues. Yet, the pattern of results was not entirely identical in both experiments. First, the consonant bias in Delle Luche *et al.* (2014) was found in the listeners' response times, but not in the accuracy of their responses. This difference is presumably due to ceiling effects with intact speech (mean accuracy: 95.2%). Fine spectro-temporal degradation of the stimuli in the present experiment lowered mean accuracy to 65.7% (which was still significantly above chance), which in turn allowed the observation of the consonant bias on that measure. Second, both Delle Luche *et al.* (2014) and the present experiment find a significant interaction between Prime type and Syllabic Structure, with evidence of a consonant bias only in targets with a VCVC structure.

Yet, [Delle Luche et al. \(2014\)](#) report faster response times in targets with C-related primes, while in the present experiment we observe slower response times in targets with V-related primes, i.e., those in which consonants are not preserved. Despite this difference, both patterns reveal an advantage of consonants in participants' lexical processing. Third, [Delle Luche et al. \(2014\)](#) additionally found evidence of a rhyme bias yielding a facilitatory effect of vowels (i.e., a vocalic bias) in CVCV tokens. Rhymes consist of a word's stressed vowel and subsequent phonemes. Although French does not have lexical stress, word endings are marked by lengthening, such that in CVCV tokens, rhymes comprise only the final vowel, which is shared with their V-related primes. [Delle Luche et al. \(2014\)](#) found that while a consonant bias is clearly found in the VCVC targets, the rhyme bias is stronger than the consonant bias in the CVCV targets, showing a vowel advantage. The absence of a vowel advantage in the present experiment provides supporting evidence to the proposal that fine spectro-temporal degradations hinder vowel recognition more than consonant recognition ([Shannon et al., 1995](#); [Xu et al., 2005](#)).

Accuracy analysis revealed a facilitatory effect of VCVC structures, as participants were more accurate when presented with VCVC as compared with CVCV targets. This advantage matches the findings of [Delle Luche et al. \(2014\)](#), who report faster response times to VCVC words. The advantage for vowel-initial tokens does therefore not originate from the degradation of the signal. [Delle Luche et al. \(2014\)](#) do not provide a potential explanation for this advantage. We put forward a tentative explanation for it. Word onsets have a special status in spoken word recognition ([Vitevitch, 2002](#)), and vowel recognition is more robust than consonant recognition both in intact speech as well as in degraded conditions, such as in speech in noise ([Cutler et al., 2004](#); [Meyer et al., 2010](#)). The combination of these two factors, added to the fact that only 20% of words in French begin with a vowel [29 119 out of 142 694 entries in the database of French words LEXIQUE, version 3.83 ([New et al., 2001](#))], could explain the facilitatory effect found in VCVC as compared with CVCV tokens.

While spectral degradation primarily affects vowel identification, removing the faster AM cues from the speech signal has been shown to particularly affect consonant recognition ([Xu et al., 2005](#)). In the supplementary material¹ we report the results of a study in which we assessed the cumulative effect of temporal degradation on the robustness of the consonant bias in word recognition, by further degrading the AM cues of the stimuli used in the present experiment. Specifically, we filtered out the fast AM components [>16 Hz, cutoff frequency used by [Shannon et al. \(1995\)](#) and [Xu et al. \(2005\)](#)]. As predicted, there was no evidence of a consonant bias in the participants' accuracy or in their response times. Note, however, that there was an acute drop in participants' mean accuracy to chance levels (48.9%), which prevent us from drawing any definitive conclusion. Future research will seek to determine the impact of temporal degradation and the effect of number of bands on the consonant bias in word recognition.

5. Conclusion

Consonants facilitate adult listeners' lexical processing to a greater extent than vowels in most languages, and this consonant bias is argued to play an important role in lexical processing and acquisition. In an auditory priming lexical decision experiment in French, a language with a clear consonant bias, we show that this consonant bias is also observed when only AM cues are preserved in eight broad spectral bands (thus, reducing fine spectro-temporal modulations of speech), a sound condition mimicking cochlear implant processors. This result suggests that current cochlear implants might allow users to rely on similar strategies as non CI users, at least for this aspect of lexical processing.

Acknowledgments

We wish to thank Marielle Hababou Bernson for her help running a subset of the participants and Julián Villegas for his help with statistical analysis. This research was supported by the Agence Nationale de la Recherche (ANR), France, under Grant No. ANR-17-CE28-0008 DESIN awarded to L.C., the ANR's French Investissements d'Avenir—Labex EFL Program under Grant No. ANR-10-LABX-0083 awarded to T.N., and the Spanish Ministry of Science and Innovation under Grant No. PID2019-105100RJ-I00 and the Basque Foundation for Science Ikerbasque awarded to I.d.I.C.P.

References and links

¹See supplementary material at <https://www.scitation.org/doi/suppl/10.1121/10.0019576> for the dataset (SuppPub1.xls) and an additional study (SuppPub2.pdf).

² $\text{glmer}(\text{Accuracy} \sim \text{PrimeType} + \text{SyllStructure} + (1 | \text{Subject}) + (1 | \text{Item}), \text{data} = \text{data}, \text{family} = \text{binomial}, \text{control} = \text{glmerControl}(\text{optimizer} = \text{"bobyqa"}, \text{optCtrl} = \text{list}(\text{maxfun} = 100000)))$.

³ $\text{lmer}(\text{RTs} \sim \text{PrimeType} * \text{SyllStructure} + (1 | \text{Subject}) + (1 | \text{Item}), \text{data} = \text{data}, \text{REML} = \text{F})$

Ardoint, M., and Lorenzi, C. (2010). "Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues," *Hear. Res.* **260**, 89–95.

Arjmandi, M. K., Jahn, K. N., and Arenberg, J. G. (2022). "Single-channel focused thresholds relate to vowel identification in pediatric and adult cochlear implant listeners," *Trends Hear.* **26**, 233121652210953.

Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (2005). "Linguistic constraints on statistical computations the role of consonants and vowels in continuous speech processing," *Psychol. Sci.* **16**, 451–459.

Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoncini, J., and Lorenzi, C. (2014). "The role of spectro-temporal fine structure cues in lexical-tone discrimination for French and Mandarin listeners," *J. Acoust. Soc. Am.* **136**, 877–882.

- Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (2006). "Acquiring an artificial lexicon: Segment type and order information in early lexical entries," *J. Mem. Lang.* **54**, 1–19.
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). "Patterns of English phoneme confusions by native and non-native listeners," *J. Acoust. Soc. Am.* **116**, 3668–3678.
- Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., and Nazzi, T. (2014). "Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study," *J. Mem. Lang.* **72**, 1–15.
- DiNino, M., Wright, R. A., Winn, M. B., and Bierer, J. A. (2016). "Vowel and consonant confusions from spectrally manipulated stimuli designed to simulate poor cochlear implant electrode-neuron interfaces," *J. Acoust. Soc. Am.* **140**, 4404–4418.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Escudero, P., Mulak, K. E., and Vlach, H. A. (2016). "Cross-situational learning of minimal word pairs," *Cogn. Sci.* **40**, 455–465.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fu, Q.-J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Hartig, F. (2022). "DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models," R Package version 0.4.5, <https://CRAN.R-project.org/package=DHARMA> (Last viewed May 17, 2023).
- Havy, M., Serres, J., and Nazzi, T. (2014). "A consonant/vowel asymmetry in word-form processing: Evidence in childhood and in adulthood," *Lang. Speech* **57**, 254–281.
- Kates, J. M. (2011). "Spectro-temporal envelope changes caused by temporal fine structure modification," *J. Acoust. Soc. Am.* **129**, 3981–3990.
- Meyer, B. T., Jürgens, T., Wesker, T., Brand, T., and Kollmeier, B. (2010). "Human phoneme recognition depending on speech-intrinsic variability," *J. Acoust. Soc. Am.* **128**, 3126–3141.
- Moore, B. C. (2004). *An Introduction to the Psychology of Hearing* (Academic, San Diego), Vol. 4.
- Moore, B. C. (2007). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (Wiley, New York).
- Moore, B. C. J. (2003). "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," *Speech Commun.* **41**, 81–91.
- Nazzi, T., and Cutler, A. (2019). "How consonants and vowels shape spoken-language recognition," *Annu. Rev. Linguist.* **5**, 25–47.
- Nespor, M., Peña, M., and Mehler, J. (2003). "On the different roles of vowels and consonants in speech processing and language acquisition," *Ling. Linguaggio* **2**, 203–230.
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). "Une base de données lexicales du Français contemporain sur internet: LEXIQUE™/A lexical database for contemporary French: LEXIQUE™," *Ann. Psych.* **101**, 447–462.
- Nishibayashi, L.-L., and Nazzi, T. (2016). "Vowels, then consonants: Early bias switch in recognizing segmented word forms," *Cognition* **155**, 188–203.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/> (Last viewed May 17, 2023).
- Shannon, R. V. (2012). "Advances in auditory prostheses," *Curr. Opin. Neurol.* **25**, 61–66.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). "Chimaeric sounds reveal dichotomies in auditory perception," *Nature* **416**, 87–90.
- Van Ooijen, B. (1996). "Vowel mutability and lexical selection in English: Evidence from a word reconstruction task," *Mem. Cogn.* **24**, 573–583.
- Vitevitch, M. S. (2002). "Influence of onset density on spoken-word recognition," *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 270–278.
- Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). "Relative contributions of spectral and temporal cues for phoneme recognition," *J. Acoust. Soc. Am.* **117**, 3255–3267.
- Xu, L., and Zheng, Y. (2007). "Spectral and temporal cues for phoneme recognition in noise," *J. Acoust. Soc. Am.* **122**, 1758–1764.
- Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargava, A., Wei, C., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2293–2298.