Journal of Phonetics 62 (2017) 1–11

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Assessing the distinctiveness of phonological features in word recognition: Prelexical and lexical influences

Alexander Martin *, Sharon Peperkamp

Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Études Cognitives, École Normale Supérieure – PSL Research University, 29 rue d'Ulm, 75005 Paris, France

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 17 January 2017
Accepted 24 January 2017

Keywords:
Word recognition
Phonological features
Perceptual similarity
Lexicon
Functional load

ABSTRACT

Phonological features have been shown to differ from one another in their perceptual weight during word recognition. Here, we examine two possible sources of these asymmetries: bottom-up acoustic perception (some featural contrasts are acoustically more different than others), and top-down lexical knowledge (some contrasts are used more to distinguish words in the lexicon). We focus on French nouns, in which voicing mispronunciations are perceived as closer to canonical pronunciations than both place and manner mispronunciations, indicating that voicing is less important than place and manner for distinguishing words from one another. We find that this result can be accounted for by coalescing the two sources of bias. First, using a prelexical discrimination paradigm, we show that manner contrasts have the highest baseline perceptual salience, while there is no difference between place and voicing. Second, using a novel method to compute the functional load of phonological features, we show that the place feature is most often recruited to distinguish nouns in the French lexicon, while voicing and manner are exploited equally often.

1. Introduction

What makes two words sound similar to each other? Consider the English word *pin* – /pɪn/. Intuitively, we can understand how a word like *shin* – /ʃɪn/ sounds more similar to *pin* than a word like *train* – /tɹe͡ɪn/ does. Indeed *pin* and *shin* form a minimal pair; the two words are minimally different, in that they share all but one phoneme. Yet cross-modal priming experiments have shown that a word like *bin* – /bɪn/, which also forms a minimal pair with *pin*, more strongly activates *pin* than *shin* does (e.g., Connine, Blasko, & Titone, 1993; Milberg, Blumstein, & Dworetzky, 1988). This is because the segments that distinguish *pin* from *shin* share fewer phonological features than those that distinguish *pin* from *bin*. Now consider the word *tin* – /tɪn/. Both the /t/ in *tin* and the /b/ in *bin* are one feature different from the /p/ in *pin* (a difference in place and voicing[1] respectively). Is the nature of the featural difference pertinent for the notion of similarity?

Research on lexical perception has demonstrated that featural differences in one's native language are not all perceived as equally distinct. In both English (Cole, Jakimik, & Cooper, 1978) and Dutch (Ernestus & Mak, 2004), mispronunciations have been shown to be less disruptive for word recognition (i.e., easier to recognize) if they involve a change in voicing than if they involve a change in place or in manner. This indicates that a difference in voicing is perceived as less stark than a difference in another major class feature in these languages. More recently, Martin and Peperkamp (2015) exposed French listeners to a series of auditorily- (or audiovisually-) presented nouns supposedly produced by a stroke patient. These included correctly pronounced words, mispronounced words, and non-words that did not resemble any real word. Participants were asked to press a button when they recognized a word – whether it was correctly pronounced or mispronounced – and report it. All mispronunciations involved a change in one of the major class features: voicing, manner, or place on a word-initial obstruent. The results from the audio-only version of that experiment, reported as the proportion of correctly identified mispronounced words, are reproduced in Fig. 1.[2] Similar

* Corresponding author.
E-mail address: alxndr.martin@gmail.com (A. Martin).

[1] Note that throughout this paper we will refer to any two-way laryngeal contrast as "voicing", although the phonetic realization of this contrast may vary across languages.

[2] The results were not significantly different by modality.

http://dx.doi.org/10.1016/j.wocn.2017.01.007
0095-4470/

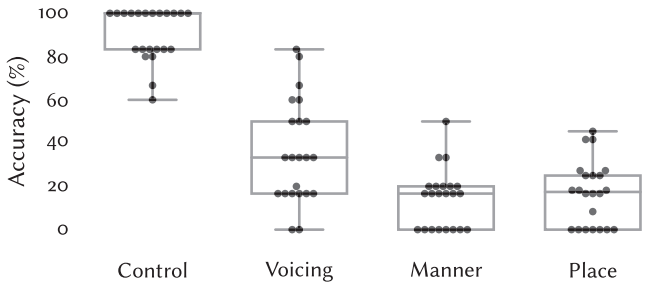

Fig. 1. Boxplot of participant means from the audio-only version of the mispronunciation detection task reported in Martin and Peperkamp (2015) by condition. The central line in the boxplot represents the median; the space between the central line and the bottom or top of the box represents the second and third quartile spread; and the distance from the bottom or top of the box to the tip of the whiskers represents the first and fourth quartile spread. In the dotplots, each dot represents an individual's score.

to the previous findings for English and Dutch (Cole et al., 1978; Ernestus & Mak, 2004), words with a voicing mispronunciation were more likely to be recognized than those with a manner or a place mispronunciation. For example, the word *sommet*, /sɔmɛ/ – "summit" was more likely to be recognized when it was presented as /zɔmɛ/, with a mispronunciation in voicing, than when it was presented as /fɔmɛ/ or /tɔmɛ/ (a place or manner mispronunciation, respectively). Thus, the voicing feature's role in contrasting words from one another is perceived as different than that of the other features.

The sources of this asymmetry remain unclear, however, and could be due to a number of factors. Most obvious is the acoustic proximity of the sounds being considered. Some sounds are acoustically closer, and will thus be perceived as more similar than other, distant, sounds. A further source of bias is language-specific knowledge. Listeners may use knowledge of their native language for the purposes of efficient word recognition. That is, they may preferentially attend to cues associated with featural contrasts which are more informative in their language. Indeed, listeners are influenced both by acoustic information and by language-specific knowledge (Ernestus & Mak, 2004; Johnson & Babel, 2010). Ernestus and Mak (2004), for example, argued that Dutch listeners are influenced by a process of initial fricative devoicing in their language, which renders voicing information on these segments uninformative. This would explain why these listeners ignore voicing mispronunciations more often than manner mispronunciations in a lexical decision task. Similarly, Johnson and Babel (2010) found language-specific influence using a similarity judgment task. They had native English- and Dutch-speaking participants rate the similarity of pairs of VCV non-words containing English fricatives, and showed that Dutch listeners rated [s], [ʃ], and [θ] as more similar to each other than English listeners did. They argued that this is due to the phonological status of these sounds in the respective languages. While all three sounds are distinctive in English, [ʃ] and [θ] are not phonologically distinctive in Dutch; the former is a contextual allophone of /s/ and the latter does not occur at all. However, in an AX discrimination experiment, Dutch listeners' response times were not shown to differ from English listeners'; both groups discriminated the same pairs of sounds equally well. The authors argued that their discrimination task reveals low-level acoustic differences between the stimuli, with some of the contrasts yielding longer response times because of their acoustic proximity (e.g., [f]∼[θ] and [h]∼[x]), regardless of the native language of the listener, while their similarity judgment task reveals language-specific influences, with Dutch listeners being perturbed by the absence of [θ] and [ʃ] as phonemes in their native language.

Note, though, that this reasoning does not explain why in English and French, voicing mispronunciations are also harder to detect (Cole et al., 1978; Martin & Peperkamp, 2015), because the voicing feature is fully distinctive in these languages (voicing contrasts can be neutralized in English and French but never word-initially). These results do not necessarily imply that listeners are not influenced by lexical patterns during word recognition. Indeed, following, inter alia, Hall (2013), we argue that a more gradient understanding of "distinctiveness" is necessary to properly address this issue. If, for example, there were *fewer* voicing minimal pairs than place and manner minimal pairs in English and French, this could explain why words presented with voicing mispronunciations were perceived as closer to the target word. Here, we further explore gradient distinctiveness using a combination of experimental and computational techniques.

The specific aim of our research is to disentangle low-level, prelexical influences from top-down, lexical ones in word recognition. To this end, we take French obstruents as a case study, allowing for a direct comparison with the results on lexical perception from the mispronunciation detection task reported in Martin and Peperkamp (2015), which we take as our starting point. Building on those results, we start off with an examination of the way phonetic differences between features are perceived outside of lexical context, using a prelexical discrimination task. We then examine the French lexicon by measuring the *functional load* of various feature contrasts as a proxy for the lexical knowledge shared by speakers of French. This allows us to understand if there are asymmetries in the usage of these different features, even though they are not affected by any phonological process. Finally, we compare our results with the word recognition results reported in Martin and Peperkamp (2015), and propose that the relative weight of phonological features during word recognition is determined jointly by the role of these features in both bottom-up acoustic perception and top-down lexical knowledge.

2. Prelexical perception

The perceptual similarity of speech sounds has been investigated for decades, focusing mostly on the effects of different types of noise on perceptual confusion (e.g., Bell, Dirks, & Carterette, 1989; Cutler, Weber, Smits, & Cooper, 2004; Miller & Nicely, 1955; Weber & Smits, 2003). For instance, Miller and Nicely (1955) presented a series of English syllables embedded in different kinds of noise (including low-pass filtering and white noise) at various signal-to-noise ratios (SNR) and asked participants to report what consonant the syllable began with. They found that *place* of articulation was more likely to be confused than *voicing*, across consonants and across different SNRs. While this line of research is important for understanding speech perception in noisy conditions, it cannot provide us with an accurate *baseline* of perceptual similarity of speech sounds, because noise affects individual

features differentially (for discussion, see Bell et al., 1989; Cutler et al., 2004).

Some studies have addressed the question of perceptual similarity in silence. However, in the absence of noise, listeners are exceedingly good at identifying sounds in their native language, hence observing differences between different types of contrasts is difficult. For instance, Plauché, Delogu, and Ohala (1997) found that Spanish listeners correctly identify the initial consonant of the syllables /pi/, /ti/, and /ki/ in 95.4% of trials; only when the stimuli were artificially manipulated did participants begin to confuse them. Wang and Bilger (1973) similarly reported ceiling performance in syllable identification, unless the stimuli were presented at low volume. Finally, in a study on the perceptual confusability of currently merging vowels in Parisian French, Hall and Hume (2013) obtained identification scores at ceiling for control, non-merging vowels (an average of 93%).

Thus, in ideal listening conditions, native sounds are reliably identified. How, then, can we measure perceptual similarity without degrading the acoustic signal? Many studies have used explicit similarity judgments: participants are asked to compare two pairs of non-words that each differ in one segment, and explicitly state which pair they find to be more similar. This methodology has been used, for example, to explore the role of feature differences in word similarity (Bailey & Hahn, 2005; Hahn & Bailey, 2005). These studies have revealed that the more phonological features the two differing sounds share, the more likely the non-words are to be judged as similar, in line with research from the lexical perception literature mentioned above (Connine et al., 1993). Although this same methodology could be applied to our current research question, related to the nature of the feature differences themselves, the fact that explicit similarity judgments require participants to metalinguistically reflect on the stimuli makes this paradigm less than ideal. Johnson and Babel (2010) compared their similarity judgment results with those from a discrimination task and showed that language-specific influences were more readily reflected in the explicit judgment task. Results of their discrimination task, they argue, were more driven by low-level acoustic perception. In the present study, we will therefore likewise use a discrimination task to test perceptual similarity in listeners' native language.

Discrimination tasks are routinely used to assess the perception of non-native and second language sound contrasts. Here, we design an ABX discrimination paradigm that aims at avoiding ceiling effects when used with native listeners. First, the stimuli are produced by multiple synthesized speakers, two male and one female, thus augmenting acoustic variability between tokens. Second, we use long (trisyllabic) stimuli, thereby increasing working memory load. Third, and most importantly, in each AB pair for which a given consonantal contrast is being tested, only one vowel and two consonants are used. For instance, for a trial with the contrast /p/-/b/, A and B might be /pa**b**aba/ and /pa**p**aba/. The fact that the crucial consonants occur multiple times should make the task particularly difficult. Note that in the given example the difference between the two items lies in the second syllable (in bold); by randomly varying this position across trials we make the task even harder, since participants cannot predict where the crucial difference will appear in a given trial.

We use this methodology to test the perceptual similarity of French obstruents that differ in only one phonological feature.

2.1. Methods

We follow Martin and Peperkamp (2015) in studying "one-feature" (major class features) obstruent contrasts in French. French has twelve obstruents that are defined by these three featural contrasts: voicing (voiced vs voiceless), manner (stop vs fricative), and place (labial vs coronal vs post-coronal). This yields twenty-four one-feature contrasts (Table 1).

2.1.1. Stimuli

For each of the 24 one-feature consonant contrasts, we constructed 18 non-word items, for a total of 432 items. Each item had the structure CVCVCV. Its vowels were identical and were drawn from the French point vowels (i.e., /a/, /i/, and /u/). The consonants were the ones from the contrast under consideration; one of them occurred once and the other one twice. By way of example, the complete set of 18 items for the contrast /p/∼/b/ is shown in Table 2. Note that it is either one or the other making up the contrast, and that it occurs either in the first, the second, or the third syllable.

All stimuli were synthesized using the Apple Say program's diphone synthesizer; each was produced by three of the European French voices: two male (*Thomas* and *Sébastien*) and one female (*Virginie*). This gave us a total of 1296 unique tokens (432 items × 3 voices). We chose to use synthesized stimuli given the large number of items and their tongue twister-like construction. The stimuli had a mean duration of 727 ms (±85 ms), and sounded relatively natural. They may be downloaded from the first author's website.

2.1.2. Procedure

A total of 1728 unique trials were created by combining the stimuli into ABA, ABB, BAA, and BAB trials. These were counterbalanced into twelve 144-trial-long lists (with each participant seeing only one list), such that each list contained a total of six trials for each of the twenty-four obstruent contrasts, including two trials for each vowel /a, i, u/.

Participants sat in front of a computer screen in a sound-attenuated room while stimuli were played binaurally through a headset. They read instructions presented on screen that described the ABX paradigm: In every trial, the first two non-words they heard would be different and the third one would always correspond to either the first or second one; they should indicate whether the third stimulus matched the first or the second; and they should give their response by pressing one of two buttons on a response box.

In each trial, participants heard a sequence of three stimuli, each produced by a different voice, with an ISI of 300 ms. Thus, X was acoustically different from both A and B. The position of the difference between A and B changed from one trial to another, making it impossible for participants to know where specifically to attend upon hearing the A stimulus of a given trial. The position of the difference was counterbalanced across trials. Voice order was randomized. For example, a participant might hear /pa**p**aba/$_{Virginie}$ – /pa**b**aba/$_{Thomas}$ – /papaba/$_{Sébastien}$, to which they should respond 'A'. In another trial, they might hear /**p**utupu/$_{Thomas}$ – /**t**utupu/$_{Virginie}$ –

Table 1
French obstruents.

	plosive		*fricative*	
	voiceless	voiced	voiceless	voiced
Labial	p	b	f	v
Coronal	t	d	s	z
Post-coronal	k	g	ʃ	ʒ

Table 2
Non-word items used for the contrast /p/∼/b/.

	/a/	/i/	/u/
/p/ × 1, /b/ × 2	/pababa/	/pibibi/	/pububu/
	/bapaba/	/bipibi/	/bupubu/
	/babapa/	/bibipi/	/bubupu/
/p/ × 2, /b/ × 1	/bapapa/	/bipipi/	/bupupu/
	/pabapa/	/pibipi/	/pubupu/
	/papaba/	/pipibi/	/pupubu/

/tutupu/$_{Sébastien}$, to which they should respond 'B'. Following a given trial, the next was presented 1250 ms after the participant had given a response, or after a timeout of 2000 ms after stimulus offset, whichever came first. If participants failed to give a response before the timeout, this was counted as an error. No feedback was given to participants during the experiment.

The experiment lasted about 20 min.

2.1.3. Participants

Forty-eight native speakers of French participated (34 women, 14 men) and were randomly assigned to one of the twelve counterbalanced lists. They were aged between 18 and 35 (mean: 25.1). None of them reported any history of hearing problems.

2.2. Results

The mean accuracy scores, in addition to the response times for correct trials (measured from the onset of the third stimulus) per phonological feature, are shown in Fig. 2. All analyses were performed using mixed-effects models in R (Bates, Maechler, Bolker, & Walker, 2014). Accuracy was analyzed with a logistic mixed-effects regression model; the log response times were analyzed using a linear mixed-effects regression model.

Both models included random intercepts for Participant and Contrast (the phoneme pair). Original models were constructed that also included a random slope for Feature by Participant, but as these models did not converge, we simplified the random effects structure. The final models included the factor Feature (voicing vs manner vs place) and random intercepts for Participant and Contrast. We took one of the values of the Feature factor as intercept and then releveled the data to assess the three-way comparison. We started by taking manner as the model intercept, comparing manner to voicing and manner to place. We then took place as the intercept, which allowed us to compare it to voicing (and redundantly to manner). The implementation of the analysis is available on the first author's website.

In both the accuracy scores and the response times, manner was found to be significantly different from voicing (accuracy: $\beta = -0.50$, $SE = 0.15$, $z = -3.36$, $p < 0.001$; RT: $\beta = 0.06$, $SE = 0.03$, $t = 2.30$),[3] while manner was found to be different from place in accuracy ($\beta = -0.43$, $SE = 0.13$, $z = -3.30$, $p < 0.001$), but only marginally so in response times ($\beta = 0.04$, $SE = 0.02$, $t = 1.83$). Voicing did not differ from place in accuracy ($\beta = -0.07$, $SE = 0.13$, $z < 1$) or response times ($\beta = 0.02$, $SE = 0.02$, $t < 1$).

These results indicate that manner contrasts yielded more correct responses than both place and voicing contrasts, and that on correct trials, manner contrasts yielded faster responses than voicing contrasts. Furthermore, place and voicing contrasts did not yield different results from each other in either accuracy or RT. In other words, manner contrasts are generally perceived as being more distinct than the other two types of contrast.

2.3. Discussion

In this experiment, we tested the perceptual similarity of French obstruents that differ from one another in only one phonological feature, using an ABX discrimination task. We found significant differences in both accuracy and response times for manner contrasts compared to place and voicing contrasts. That is, participants were both more accurate and faster to discriminate obstruents differing in manner than obstruents differing in place or voicing. This result indicates that obstruents differing in manner are on average perceived as being more distinct than those differing in place or voicing, even though all of the contrasts we tested are phonologically distinctive in French.

From an acoustic point of view this makes sense, given that the manner contrasts we tested differentiated stops from fricatives. That is, stops are characterized by a period of silence followed by a burst, whereas fricatives involve aperiodic noise throughout their production; intuitively, the difference between a period of silence and a period of noise is easy to perceive. Unsurprisingly, our results do not mirror those mentioned above concerning the perception of speech in noisy conditions. Indeed, the difference between a period of silence (stops) and a period of noise (fricatives) can be easily masked when additional noise is superimposed onto the stimuli, thereby diminishing perceptual differences between the two. The presence or absence of low-frequency periodicity (voiced vs voiceless sounds respectively), and the formant transitions associated with distinguishing different places of articulation may resist such noise manipulation better. The present results, then, give credence to our claim that the study of speech in noise cannot provide us with an accurate baseline of perceptual similarity.

We verified that stark acoustic differences drive effects in the prelexical task by performing acoustic analyses on our stimuli. We used the spectral package in Python (Versteegh, 2015) to extract forty-dimensional Mel filterbanks coefficients with a cubic-root compression for each stimulus; these acous-

[3] The lme4 package does not provide *p* values for linear models; a *t* value greater than 2 is usually considered to be significant. Given our experimental design (specifically the number of participants and the number of items), this method has been shown to have the lowest Type I error of the common methods for assessing significance of mixed-effects models (Luke, 2016).

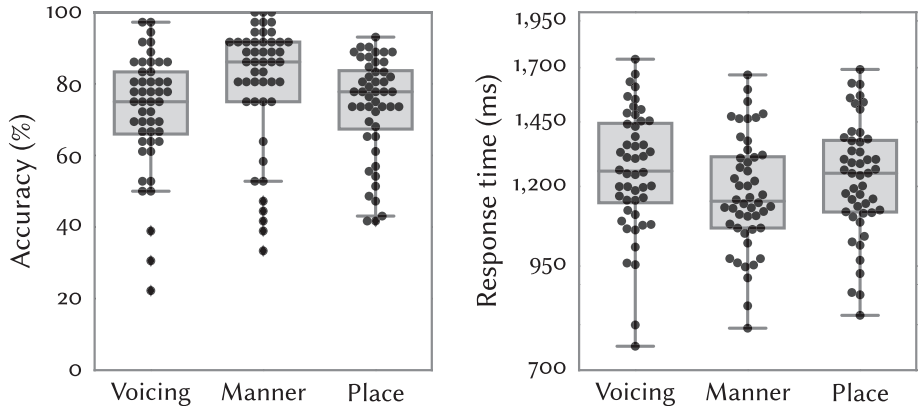

Fig. 2. Box- and dotplots of participant means of accuracy (left) and response times (on a log scale) on correct trials (right) by feature.

tic features are meant to roughly reflect the way the sounds are represented at the level of the cochlea. We then used the ABXpy package (Schatz, 2016; Schatz et al., 2013) to model predicted performance in an ABX task taking into account only the acoustic properties of the stimuli. This model uses Dynamic Time Warping (Rabiner & Juang, 1993) to measure the acoustic distance between the stimuli A and X on the one hand and B and X on the other hand, in order to predict a response based on which distance is shorter (i.e., if the A-X distance is shorter than the B-X distance, the predicted response is A). While the model's overall performance is lower than that of the participants in our experiment, the pattern is exactly the same: The model performs better on manner contrasts (mean: 69.6%) than it does on voicing (mean: 61.5%) or place (mean: 60.8%) contrasts, with performance on the latter two being nearly identical.

Previous attempts to measure perceptual similarity of speech sounds in silence have yielded ceiling performance, and it is only when noise is added that stimuli begin to be confused by participants. Our paradigm, on the other hand, allowed us to measure asymmetries in processing without degrading the speech sounds. The similarity of the A, B, and X tokens (including many repetitions of each of the target sounds) made the task sufficiently difficult that we did not observe ceiling performance. This crucially provides us with a baseline of perceptual similarity of the sounds we tested, and our task could also be used to study native listeners' perception of other consonantal, vocalic, or even tonal contrasts.[4]

Returning to the question of word-level similarity, the results from this experiment are only partially in line with Martin and Peperkamp (2015)'s lexical results, according to which words with either a manner or a place mispronunciation are harder to recognize as the intended real words than words with a voicing mispronunciation. Although the present result can explain why manner mispronunciations are more disruptive for the purposes of word recognition (i.e., they are perceived as being more different from the canonical pronunciations), it does not explain why place mispronunciations are equally disruptive. We hypothesize that the latter result is due to another bias, namely lexical knowledge. Indeed, lexical effects can be observed at many levels of speech processing, including phonological judgments (e.g., Hay, Pierrehumbert, & Beckman, 2004). In the following section, we quantify the relative weight of phonological features in the lexicon, using a new measure of functional load, which we propose as a proxy for lexical knowledge.

3. Functional load

The term *functional load*, in a broad sense, refers to the amount of work a phonemic contrast does in a language to distinguish words from one another. Consider the English distinction between /θ/ and /ð/ (the "th" sounds in *think* and *that* respectively). Although these sounds are distinctive in English, they actually only disambiguate a handful of words (e.g., *ether*∼*either* in American English), and the contrast is therefore considered to have a low functional load. Compare this to the high functional load of the contrast /p/ ∼ /t/, which disambiguates a great many pairs of words (e.g., *pack*∼*tack*, *pin*∼*tin*, *cope*∼*coat*).

Functional load has been proposed as a key factor in language change (Martinet, 1955). Specifically, contrasts that have low functional load are predicted to be more likely to merge over time than contrasts that have high functional load. This hypothesis has been put to the test by, for instance, Wedel, Kaplan, and Jackson (2013), who showed that in many languages, contrasts that have low functional load are indeed more likely to merge over time. They compared two specific measures of functional load: minimal pair counts, and difference in information entropy. The first is rather straightforward. Reconsider the English examples from above: only seven pairs of words are disambiguated by the /θ/∼/ð/ contrast, compared to well over 300 for the /p/∼/t/ contrast. The second measure of functional load, based in information theory (Shannon, 1948), concerns information entropy (Hockett, 1955). This is a quantification of the uncertainty of the system, with the lexicon considered to be a complex system. The higher the functional load of a given contrast, the more "uncertainty" is removed from the system, should that contrast be excised. The measure is therefore calculated as the difference in entropy of the system with or without the contrast; the greater the difference, the higher the functional load. For a detailed mathematical description of the calculation of this measure, including its application to different levels of analysis,

[4] We actually used the experimental design reported in this study to test one-feature vowel contrasts in French as well. The results of that study are not reported here, but the data are available on the first author's website.

see Surendran and Niyogi (2003, 2006). While the entropy measure has been widely used (e.g., Hume et al., 2013; van Severen et al., 2013; Stevenson, 2015; Stokes & Surendran, 2005), Wedel, Kaplan et al. (2013) found that, overall, minimal pair counts are a more accurate predictor of sound change, and more recent work has backed up this finding (Wedel, 2015).

Here, we are interested in the functional load of phonological features (e.g., voicing), rather than of individual phonemic contrasts (e.g., /p/∼/b/). Previous research has indicated that a handful of these individual contrasts tend to be responsible for distinguishing a disproportionate amount of words from one another in the lexicon of a given language (Oh, Coupé, Marsico, & Pellegrino, 2015), but has not examined whether these contrasts concern the same phonological feature. Thus, are high functional load phoneme pairs likely to contrast in the same feature? Minimal pair counts and entropy measures have been adapted to respond to this question, but the results are slightly problematic. For minimal pairs, calculating such a score equates to summing the minimal pair counts for all contrasts in a given feature; for entropy, calculating scores for features rather than for phoneme contrasts can be done by summing the scores for all phoneme contrasts within a given feature. For instance, for both minimal pair counts and entropy, in a four-phoneme system like /p, b, t, d/, the score for place would be the summed scores for /p/∼/t/ and /b/∼/d/. Likewise, the score for voicing would be the summed scores for /p/∼/b/ and /t/∼/d/; thus one value per feature. Surendran and Niyogi (2003) report a series of values obtained by applying the entropy measure in this way cross-linguistically, but their method raises the question of how to interpret an absolute difference of, say, 0.002 or 0.009 bits.[5] This problem pertains to all the existing implementations of both the entropy and the minimal pair methods. Are observed differences meaningful, or do they simply reflect some kind of noise? Assessing the significance of a difference typically relies on inferential statistics, for which distributions of scores rather than single numbers are required. Below, we propose a new method of measuring functional load that is based on minimal pair counts but that allows for the use of inferential statistics. Our method also differs from previous ones in another aspect. Existing minimal pair counts and entropy methods provide scores of the *absolute* functional load of a given contrast within a system. They are thus affected by the individual frequency of the phonemes that form the contrast. That is, the theoretically maximum functional load of a phoneme contrast depends upon the frequency of the phonemes in question. This issue has been reported in a previous study on the correlation between functional load and perceptual similarity (Hall, 2009). It is also problematic for our present purpose of measuring the functional load of phonological features, as the functional load of a feature (e.g., voicing) is a function of the functional load of the relevant phoneme contrasts (e.g., /p/∼/b/, /t/∼/d/, . . .). We ask here to what extent each feature is used, all else being equal. We thus propose a *relative* measure of functional load, which abstracts away from individual phoneme frequency.

In the same vein, our measure abstracts away from phonotactics, that is, the constraints governing the combination of sounds in a language, which may make a given contrast impossible in certain positions. In French, for instance, the initial cluster /tl/ is not permitted (Dell, 1995). It is therefore impossible for /pl/-initial words, such as /plɥi/ (*pluie*, “rain”), to be changed into another word by replacing /p/ with /t/. We consider this to be uninformative regarding the distinctive weight of /p/∼/t/ (or of the place feature for that matter). We specifically place phonotactic knowledge on a different level from lexical structure. Our question should instead be framed as: When /p/ *can* contrast with /t/, does it? And how does the frequency with which it does compare to the frequency with which /p/ contrasts with /b/ when replacing /p/ by /b/ is phonotactically legal?

Below, we detail our new method.

3.1. A new measure of functional load

Given the previous success of minimal pair counts in assessing the functional load hypothesis for sound change (Wedel, Kaplan et al., 2013), our measure of the functional load of phonological features is based on minimal pair counts. It consists of an observed-over-expected ratio (henceforth O/E ratio), as defined in Eq. (1).

$$O/E\,ratio_{i,j} = \log\left(\frac{o_{i,j}}{e_{i,j}}\right) \tag{1}$$

For each phoneme *i*, and each feature *j*, an observed-over-expected score is calculated, where *e* represents the number of possible (or expected) minimal pairs and *o* the number of observed minimal pairs for that phoneme in that feature. This function is iterated over the lexicon for each feature, and each phoneme. For example, consider the French phoneme /p/ and the place feature (here, specifically the change from /p/ to /t/, although the change from /p/ to /k/ would also need to be included). Upon encountering the word /po/ (*peau*, “skin”), a minimal pair is theoretically possible (i.e., a change in place on the segment /p/ yields the phonotactically legal word /to/). We consider that if the lexicon maximally exploited all contrasts, then we should *expect* to find a minimal pair between /po/ and /to/. This is precisely what the value of *e* is meant to represent. Thus $e_{/p/,\mathrm{PLACE}} = 1$. Furthermore, the word /to/ does exist (*taux*, “amount”). Thus $o_{/p/,\mathrm{PLACE}} = 1$. If we next consider a case such as /pjɛʒ/ (*piège*, “trap”), we observe that the theoretical minimal pair it would form with a place change is possible (i.e., /tjɛʒ/ is phonotactically legal). Thus $e_{/p/,\mathrm{PLACE}} = 2$ now (the scores are cumulative as we iterate over the lexicon). However, this word does not actually exist in French, and $o_{/p/,\mathrm{PLACE}}$ therefore remains at 1. Now if we consider the case of /plɥi/ (*pluie*, “rain”), we know that the theoretical minimal pair it would form from a place change is not possible (i.e., /tlɥi/ is ruled out by the French phonotactic constraint that words cannot begin with /tl/), thus $e_{/p/,\mathrm{PLACE}}$ remains at 2. Of course, if a minimal pair is not possible, it will not be observed, and indeed $o_{/p/,\mathrm{PLACE}}$ remains at 1. This process is repeated over the entire lexicon for each combination of a phoneme and a feature (so we might next consider manner minimal pairs for the sound /p/, or place minimal pairs for the sound /t/, until all combinations were exhausted), yielding distributions of scores for each feature. These scores are log-transformed to ensure they follow a normal distribution; thus a score of zero represents the highest possible functional load. That is, if every possible minimal pair

[5] Entropy is measured in “bits” of information.

were attested, the score for that phoneme and that feature would be zero. Additionally, if no minimal pairs are observed at all, the score will be $-\infty$. Note further that the operation is performed over lemma forms, not over the unique set of phonological forms, so if two lemmata have the same phonological form (i.e., they are homophones), they are both counted.

3.2. Applying O/E ratios to French nouns, and comparison with entropy

As our current question is about exploring different sources of influence on word recognition, so as to understand asymmetries reported in the literature, we focused our analysis on French nouns, the class of words tested by Martin and Peperkamp (2015). Using the Lexique database (New, Pallier, Ferrand, & Matos, 2001), we calculated the functional load of each phonological feature for each obstruent, using lemma forms as reported in Lexique. We chose the lemma forms, as functional load calculated on lemma rather than on surface forms is a better predictor of sound change (Wedel, Jackson, & Kaplan, 2013). The lemma forms of French nouns are simply the singular. Note that very few words in French are marked phonologically in the plural form (e.g., *journal* /ʒuʁnal/ – *journaux* /ʒuʁno/,"newspaper(s)"); these words would be considered only in the singular. We followed the method detailed above, respecting French phonotactic constraints. Unlike minimal pair counts and entropy differences, which are fairly straightforward to measure, our proposition requires knowledge of the language's phonotactic constraints, and further depends on their interpretation. For instance, while /tl/ is universally rejected as a possible onset in French, /pn/, which is a rare onset occurring exclusively in words of Greek origin (e.g., /pnø/ – *pneu*, "tire"), may or may not be accepted, depending on the speaker.[6] For the present purposes, we considered possible clusters those described to be well-formed according to Dell (1995), plus some of those described in the same study as rare, such as /sm/ as in *smiley* or /sn/ as in *snob* (for an exhaustive list, see Appendix). We included only rare clusters that were deemed acceptable during an informal survey. We calculated the possible minimal pairs for obstruents in all positions (i.e., not just word-initially). The results of this calculation are shown in Fig. 3. The distributions represented are the scores of the twelve phonemes for each feature. For voicing, and manner, each phoneme is involved in one comparison (e.g., /p/ vs. /b/, or /p/ vs. /f/), while for place, each phoneme is involved in two comparisons (e.g., /p/ vs. /t/ *and* /p/ vs. /k/).

We performed a one-way ANOVA on the distributions of scores obtained using our functional load method. It should be noted that instead of sampling a distribution from a population, we are indeed sampling the *entire* distribution. That is, we include all of the contrasts that each feature involves within the subset of sounds under consideration. A significant difference was observed across the phonological features ($F = 9.11$, $p < 0.001$). Post-hoc analyses using the Tukey HSD test showed that it was the place feature that was significantly different from manner ($p < 0.001$) and marginally significantly different from voicing ($p = 0.068$), but no difference was found between manner and voicing ($p > 0.05$). This indicates that the place feature has a higher functional load in French nouns than the other two tested features.

Next, we compared these results to ones obtained using a measure of the difference in entropy (Surendran & Niyogi, 2003, 2006). Although the entropy measure for features is calculated by summing the difference in entropy for each contrast within that feature (recall the mini-language with just /p, b, t, d/ described above, where the difference in entropy for voicing is equal to the differences for /p/∼/b/ and /t/∼/d/ combined), it is possible, for the purposes of performing inferential statistics, to consider each contrast as one data point in a distribution. The functional load of voicing, for example, then becomes a vector of entropy differences (in our example, {/p/∼/b/, /t/∼/d/}), allowing for statistical comparison across the features. We used the Phonological CorpusTools kit (Hall, Allen, Fry, Mackie, & McAuliffe, 2015b) to calculate the differences in entropy within French nouns extracted from the LEXIQUE database; the results can be seen in Fig. 4. The distributions represented are made up of every contrast in that dimension (e.g., for voicing /b/∼/p/, /z/∼/s/, etc.).

Note that place has a seemingly higher functional load (mean = 0.0140) than the other two features (manner: mean = 0.0133; voicing: mean = 0.0130), just as with our O/E ratio measure. A one-way ANOVA, however, revealed no significant difference amongst the features ($F < 1$). Thus, given the available data points, place's modestly higher score appears to be uninterpretable. Of course, given that the numerical pattern is similar to what we obtain with the O/E ratio method, it is likely that there is a true effect, which is simply masked (perhaps due to frequency or phonotactic effects as described above). Indeed, the O/E ratio and entropy methods are grossly measuring the same thing (use of a contrast within the lexicon), and are in fact highly correlated (Pearson's $r = 0.73$, $p = 0.001$).

3.3. Discussion

The two traditional measures of functional load[7] (minimal pair counts and information entropy) are inappropriate for comparing the distinctiveness of phonological features. Both can be biased by individual phoneme frequency, as well as by the presence of phonotactic constraints, which we would like to separate from the question of lexical distinctiveness. Indeed the traditional methods measure the absolute functional load of contrasts, while we are interested in their relative weight. The speaker may ask, "Given the constraints of my language, how distinctive is the place contrast compared to the manner contrast?" We therefore proposed a new measure, O/E ratios, that abstracts away from these properties, and additionally provides distributions of scores, allowing the use of inferential statistics to test the significance of observed differences.

When applying our method to obstruents in French nouns, we found that the place feature has a significantly higher func-

[6] This specific example is known to vary according to region, with an epenthesized version /pən/ being preferred in the south of France.

[7] We focus here on functional load, and presume it to represent the knowledge of lexical organization shared by speakers of French. Another common lexical measure, neighborhood density, is not appropriate to answer our question about single features, since it includes minimal pairs differing in multiple features, as well as minimal pairs differing in the absence versus presence of a segment (e.g., the neighborhood of the word *peau* contains not only *beau* and other one-feature change words, but also *vaut* (a multi-feature change), *eau* (a deletion), and *pôle* (an insertion)).

Appendix

French consonant clusters considered licit in the functional load analysis.

Post-pausals (onsets)			Pre-pausals (codas)		
pʁ	pɥ	bʁw	ʁp	tʁ	kstʁ
tʁ	tɥ	blw	ʁt	kʁ	
kʁ	kɥ	dʁw	ʁk	bʁ	
bʁ	bɥ	gʁw	ʁb	dʁ	
dʁ	dɥ	glw	ʁd	gʁ	
gʁ	fɥ	fʁw	ʁg	fʁ	
fʁ	sɥ	pʁɥ	ʁf	vʁ	
vʁ	ʃɥ	plɥ	ʁv	pl	
pl	ʒɥ	tʁɥ	ʁs	kl	
kl	mɥ	kʁɥ	ʁz	bl	
bl	nɥ	bʁɥ	ʁʃ	gl	
gl	ʁɥ	dʁɥ	ʁʒ	fl	
fl	lɥ	gʁɥ	ʁm	pt	
sp	pj	fʁɥ	ʁn	ps	
st	tj	flɥ	ʁɲ	ts	
sk	kj	pʁj	ʁl	tʃ	
sf	bj	tʁj	lp	tm	
sm	dj	kʁj	lt	kt	
sn	fj	bʁj	lk	ks	
pw	vj	gʁj	lb	ʁpʁ	
tw	sj	klj	ld	ʁtʁ	
kw	ʃj	spʁ	lg	ʁkl	
bw	mj	stʁ	lf	ʁbʁ	
dw	nj	skʁ	lv	ʁsk	
fw	ʁj	spl	ls	spʁ	
vw	lj	skl	lʃ	stʁ	
sw	ps		lʒ	skl	
ʃw	tʃ		lm	ltʁ	
ʒw	pʁw		sp	lkʁ	
mw	plw		st	ktʁ	
nw	tʁw		sk	kst	
ʁw	kʁw		sm	ptʁ	
lw	klw		pʁ	ʃtʁ	

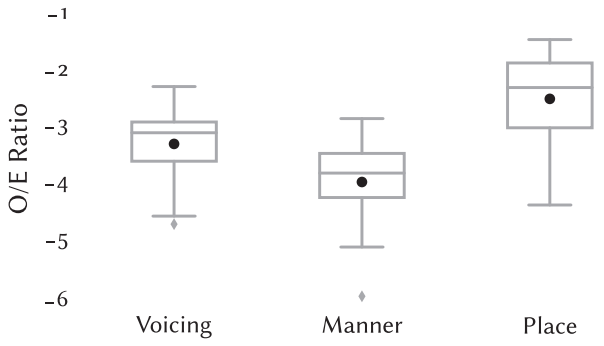

Fig. 3. Boxplots of functional load as measured with O/E ratios for French nouns. Black dots represent the means of the distributions.

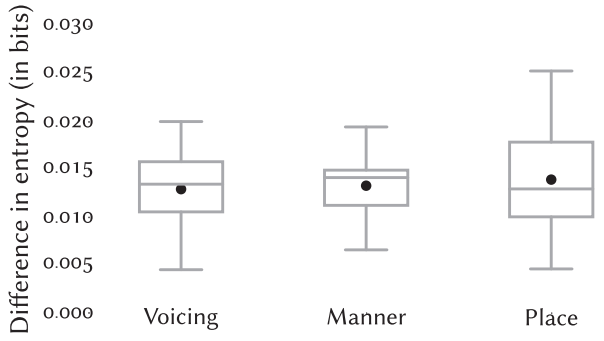

Fig. 4. Boxplots of functional load as measured by differences in entropy for French nouns. Black dots represent the means of the distributions.

tional load than the manner and voicing features. This means that in the French lexicon, nouns are more likely to be distinguished from one another by place than by manner or voicing, where possible. Because our score is based on an observed-over-expected ratio, it excludes effects due to frequency, or even due to the number of possible contrasts (French has twice as many place as manner or voicing contrasts within obstruents). Hence, any observed differences truly reflect the extent to which the contrasts are used in a distinctive way in the lexicon (at a type, rather than token, level). It would be pertinent in future applications to test to what extent weighting minimal pairs by token frequency would affect the patterns we observe.

Although our method yields results that are highly correlated with those of the difference in entropy measure, we observed significant differences within the French noun class that are not captured by the entropy measure. Moreover, it is more insightful even for results that are in line with the entropy measure. For instance, the contrasts /d/∼/z/ and /s/∼/z/ have very low functional load according to both measures. While observing this in the entropy measure alone might lead one to think that the effect is driven by the low frequency of /z/ (indeed, all four of the lowest entropy scores are contrasts involving this phoneme), the fact that we observe a similar pattern in the O/E ratios shows that even when French does use /z/, it rarely contrasts with other sounds, be it in voicing, manner, or place. This, then, shows that there is something going on *beyond* the simple distribution of sounds in the language.

The use of our method, though, requires language-specific knowledge. While we based our analysis on a phonological description of French syllable structure (Dell, 1995), a more bottom-up approach may be adopted by extracting phonotactic

rules from the corpus being studied. For the case of French, Lexique contains certain words with very rare clusters (e.g., /ft/ as in *phtaléine*); it may therefore be prudent, when using such an approach, to set a frequency threshold, including only clusters which appear a certain number of times. It further presumes a specific representation of phonotactics. The current implementation of our method considers phonotactics categorically: either a wordform is licit or it is not. However, as phonotactic acceptability is known to be gradient, and dependent on lexical statistics (e.g., Frisch, Large, & Pisoni, 2000), it may be interesting to incorporate such a notion in a future implementation of our method.

A further issue is that of phonological processes. For example, when considering a language with final devoicing, such as Dutch, the question arises as to whether to perform the calculation on underlying forms (e.g., /mud/ – "courage") or on surface forms (e.g., [mut]). If the underlying form is chosen, then /mud/ contrasts with /mut/ – "must". Of course, this is an issue for all calculations of functional load, as are the similar issues of lemma versus inflected forms and canonical versus reduced forms. For example, the English word "probably" is often produced as [pɹɑli]; as noted by Hall, Allen, Fry, Mackie, and McAuliffe (2015a), this common variant, but not the canonical pronunciation, contrasts with the word /tɹɑli/ – "trolley". Using the entropy measure, Hall et al. showed that patterns of results do not greatly differ according to whether the analysis is based on the most common pronunciation or on the canonical form. It may therefore be reasonable to likewise focus on the more abstract, underlying, level, where voicing would therefore be distinctive word-finally in a language such as Dutch. This will be an important consideration for future implementations of this measure, which did not arise in the French data, as the sounds tested are not affected by any neutralization process.

Finally, let us return to the question of relative weight of phonological features during word recognition. Based on the results from the ABX task, we argued above that prelexical perception accounts for the relative importance of manner compared to the other features, given its basis in a stark acoustic difference. The functional load results, then, provide an explanation for the relative importance of place in the results reported in Martin and Peperkamp (2015): Given that within nouns, the place feature has a higher functional load than the voicing feature, French listeners lend more importance to place than to voicing cues during word recognition, thereby making it harder to recognize a word with a changed place feature than one with a changed voicing feature.

4. General discussion

French listeners are more likely to recognize a mispronounced version of an obstruent-initial word if the mispronunciation concerns the voicing feature than if it concerns the place or manner features (Martin & Peperkamp, 2015). Thus, in French obstruents, both place and manner are more important for word recognition than voicing (at least in nouns), akin with findings in other languages (Cole et al., 1978; Ernestus & Mak, 2004). Where does this asymmetry come from? We examined two sources: prelexical acoustic perception, and lexical knowledge. In order to do so, we introduced two methodological novelties. First, we developed a version of the ABX discrimination paradigm that allows for assessing differences in the perception of native language sounds without presenting the stimuli in noise. Specifically, we increased the difficulty of the task (by using long, very similar non-words, a short ISI, and multiple voices), and showed that even among fully distinctive contrasts, some are more difficult to discriminate than others. Contrary to the two most-reported methods for assessing similarity, syllable identification and similarity judgments, our method neither yields ceiling performance, as is common in syllable identification in clear speech, nor requires participants to meta-linguistically reflect on the sounds themselves, as in explicit similarity judgments. Second, we developed a new method of measuring functional load, based on an observed-over-expected ratio of minimal pairs in the lexicon. This method is less vulnerable to language-specific tendencies that can bias traditional measures such as minimal pair counts and differences in entropy, and allows, moreover, for the use of inferential statistics to compare features amongst themselves. This makes it more appropriate for our current research question than the traditional methods of simply counting the number of minimal pairs, or of calculating difference in system entropy with and without the contrast.

Using these new methodologies, we examined the prelexical perception of obstruent features by French listeners on the one hand, and the functional load of these features in French nouns on the other hand. Results from the perception experiment showed that French listeners are better at discriminating French nonce words with obstruents that differ in manner of articulation than in place or voicing; this mirrors the fact that manner contrasts are acoustically more distant than place or voicing contrasts. Results from the functional load computation showed that within the class of French nouns, place differences are more often used to distinguish words than voicing and manner differences.

We propose, then, that French listeners are biased by both of these phenomena during word recognition. First, the strong acoustic difference between stops and fricatives makes the manner contrast easy to perceive (note that we predict this type of effect to be observable cross-linguistically). This explains why participants in Martin and Peperkamp (2015) had great difficulty recognizing words with a manner mispronunciation (for instance, replacing /v/ with /b/ disrupted recognition of the word *voleur* – "thief"). Second, the fact that their language uses the place feature more often than the other features to distinguish words leads French listeners to preferentially pay attention to place cues during word recognition. This explains why participants similarly failed to recognize words with a place mispronunciation (for instance, replacing /v/ with /ʒ/ also strongly disrupted recognition of the word *voleur*). By contrast, as voicing stands out neither in prelexical perception nor in functional load, participants had less difficulty recognizing words with a voicing mispronunciation (replacing /v/ with /f/ was less disruptive for recognition of *voleur*).

Thus, the combination of our prelexical experiment and lexical analysis can explain the word recognition results reported in Martin and Peperkamp (2015). Our results indicate that listeners are sensitive to lexical structure, and that they recruit this knowledge during word recognition. They further demonstrate that, unsurprisingly, low-level acoustic information biases listeners at multiple levels of processing (i.e., in the

higher-level lexical task as well as in the lower-level discrimination task). Our conclusion is that during word recognition, listeners' knowledge of the French lexicon coalesces with their low-level perceptual biases, yielding greater perceived distinctiveness for the manner and place compared to the voicing feature. This is in line with a vast literature on the integration of bottom-up and top-down influences (for a review, see Davis & Johnsrude, 2007).

It is, though, important to consider our specific definition of feature. We have been examining featural contrasts along the major class dimensions, without consideration of more specific features (binary or not). Of course the methodologies we have presented in this paper could be used to examine any set of features, but we do make certain assumptions that merit discussion. One major point is our consideration of the manner feature as concerning only stops and fricatives, since we restrict ourselves to obstruents. Our experiment and argumentation focus on the stark acoustic difference between these two types of sounds, which, we argue, yields the importance of the manner feature within the obstruent class. This does not necessarily transfer to other types of manner contrasts. For example, the difference between nasals and voiced stops, although a manner difference, is not automatically predicted to behave in the same way as the manner contrasts we tested here. It is entirely possible that the manner feature has a different weight relative to place and voicing when considering nasals. Our conclusions therefore must be taken within the class of sounds we tested. Future implementations of this measure might consider different types of features and different distinctions, in addition to other types of contrast (vowels, tones, etc.).

Furthermore, although our functional load measure specifically abstracts away from the number of contrasts within a certain feature by using ratios (i.e., the fact that there are twelve place contrasts but only six voicing contrasts in French obstruents is corrected for by *expecting* more place minimal pairs in the lexicon), our results do not allow us to say definitively that it is functional load that drives the importance of place for French listeners during word recognition. While we found that French uses place more than manner or voicing to distinguish nouns from one another, it is also true that place, for example, is a three-way contrast, while voicing is strictly binary. It would be interesting to focus on a language with higher dimensionality in the voicing feature (e.g., Korean or Eastern Armenian). If the number of contrasts in both the place and voicing features is the same, and speakers of such a language still pay more attention to place than to voicing in a lexical task, then our claim that it is lexical knowledge rather than knowledge of the phonological inventory that is exploited during word recognition would be bolstered.

A further consideration regarding our functional load analysis is that it focuses on nouns, and presumes that listeners track statistical information pertaining to phonological contrasts within lexical classes. This is indeed supported by various studies (Farmer, Christiansen, & Monaghan, 2006; Farmer, Monaghan, Misyak, & Christiansen, 2011; Heller & Goldrick, 2014; Strand, Simenstad, Cooperman, & Rowe, 2014). For instance, Strand et al. (2014) showed that listeners are sensitive to syntactic context during isolated word recognition. In their word identification task, accuracy was negatively affected by a measure of within-class grammatical density when syntactic category was constrained. Additionally, previous work on sound change has shown that within-category minimal pairs better predict mergers over time than across-category pairs, further suggesting that contrast may be category-sensitive (Wedel, Jackson et al., 2013). Future work on lexical influence during speech processing could further explore the role of lexical classes. In particular, we predict that any functional load differences found within the French verb class should similarly be reflected in bias during word recognition. Thus, if within verbs one feature has a higher functional load than others, we expect that mispronunciations of that feature in verbs will be more disruptive for word recognition than mispronunciations of other features.[8]

Finally, our conclusions hinge on a qualitative comparison of three results: place and manner were shown to be significantly more important for word recognition in Martin and Peperkamp (2015); the importance of manner can be explained by its acoustic saliency, as demonstrated by the prelexical experiment reported in the present study; the importance of place can be explained by its lexical status, as demonstrated by our functional load measure. However, because of the different ways that each of these results were obtained, making a *quantitative* comparison is rather difficult. For example, it may be tempting to compare the individual contrasts tested in the lexical and prelexical tasks, to examine whether performance on the manner contrasts in the prelexical task negatively correlates with performance on manner contrasts in the lexical task. This is not straightforward, though. In the ABX task, comparing X to A and B is symmetrical; if the contrast tested is /t/∼/s/, participants compare, say, /bivibi/$_X$ to both /vivibi/$_A$ and /bivibi/$_B$. By contrast, the lexical task is asymmetrical, as participants attempt to map a given mispronunciation onto an existing lexical representation. In attempting to map non-existent *boleur* to the real word *voleur*, the question can be very clearly stated: does /b/ activate /v/? Other trials using other words (e.g., non-existent *veignet* mapped to real *beignet* – "fritter") ask the reverse question. In order to directly compare the contrasts tested in the lexical task, it might be preferable to have an asymmetrical prelexical task, such as an oddball paradigm, where a deviant stimulus is compared unidirectionally to a standard.

To conclude, word recognition is a complex process that takes into account both low-level acoustic information and language-specific, phonological and lexical, knowledge. We have provided evidence that acoustic salience and lexical distinctiveness coalesce and bias listeners' weighting of phonological features. The two methodological tools that we developed can be used independently, one for assessing the prelexical discriminability of native phonemes without altering their acoustic properties, and one for assessing the relative functional load of phonological features.

[8] It would be interesting to extend the present research to French verbs, but this is less straightforward than one would hope. Recall that we used lemma forms to calculate functional load. Lexique codes the lemma form of French verbs to be the infinitive, but French verbs invariably have infinitive morphology (/-e/, /-iʁ/, or /-ʁ/) that may influence the outcome of the calculation. For example, the French verb *battre* – /batʁ/ (*to hit*) does not form a minimal pair with the possible but nonexistent form /datʁ/. The stem of the same verb, /bat/, however, does contrast with the stem /dat/ of the verb /date/ (*to be dated, old*). Thus, the choice of the lemma form impacts the functional load of, in this case, the place feature.

Acknowledgements

This work was supported by ANR-13-APPR-0012 LangLearn, ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*. We would like to thank Rory Turnbull for comments and discussion, and Thomas Schatz for his help implementing his model of the ABX task.

References

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language, 52*(3), 339–362. http://dx.doi.org/10.1016/j.jml.2004.12.003.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4.

Bell, T., Dirks, D. D., & Carterette, E. C. (1989). Interactive factors in consonant confusion patterns. *The Journal of the Acoustical Society of America, 85*(1), 339–346.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America, 64*(1), 44–56.

Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193–210.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668–3678. http://dx.doi.org/10.1121/1.1810292.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1–2), 132–147. http://dx.doi.org/10.1016/j.heares.2007.01.014.

Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua, 95*(1–3), 5–26. http://dx.doi.org/10.1016/0024-3841(95)90099-3.

Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language, 90*(1–3), 378–392. http://dx.doi.org/10.1016/S0093-934X(03)00449-8.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 103*(32), 12203–12208. http://dx.doi.org/10.1073/pnas.0602173103.

Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: Reply to Staub, Grant, Clifton, and Rayner (2009). *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(5), 1318–1325. http://dx.doi.org/10.1037/a0023063.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*(4), 481–496. http://dx.doi.org/10.1006/jmla.1999.2692.

Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition, 97*(3), 227–267. http://dx.doi.org/10.1016/j.cognition.2004.09.006.

Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. The Ohio State University.

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review, 30*(2), 215–276.

Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2015a). Calculating functional load with pronunciation variants. Stuttgart, Germany.

Hall, K. C., Allen, B., Fry, M., Mackie, S., & McAuliffe, M. (2015b). Phonological CorpusTools Retrieved from http://phonologicalcorpustools.github.io/CorpusTools/.

Hall, K. C., & Hume, E. V. (2013). Perceptual confusability of French vowels. In Proceedings of meetings on acoustics (Vol. 19). Montreal, Canada. doi: http://dx.doi.org/10.1121/1.4800615.

Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in laboratory phonology VI* (pp. 58–74). Cambridge, UK: Cambridge University Press.

Heller, J. R., & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review, 21*(6), 1576–1582. http://dx.doi.org/10.3758/s13423-014-0616-3.

Hockett, C. (1955). A manual of phonology. *International Journal of American Linguistics, 21*(4).

Hume, E., Hall, K. C., Wedel, A., Ussishkin, A., Adda-Decker, M., & Gendrot, C. (2013). Anti-markedness patterns in French epenthesis: An information-theoretic approach. In Proceedings of the 37th annual meeting of the Berkeley linguistics society (pp. 104–123).

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics, 38*(1), 127–136. http://dx.doi.org/10.1016/j.wocn.2009.11.001.

Luke, S. G. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 1–9. http://dx.doi.org/10.3758/s13428-016-0809-y (Mcmc).

Martin, A., & Peperkamp, S. (2015). Asymmetries in the exploitation of phonetic features for word recognition. *The Journal of the Acoustical Society of America, 137*(4), EL307–EL313. http://dx.doi.org/10.1121/1.4916792.

Martinet, A. (1955). Économie des changements phonétiques (Francke). Bern.

Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society, 26*(4), 305–308.

Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338–352.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). *Une base de données lexicales du français contemportain sur internet: LEXIQUE. L'année Psychologique, 101*, 447–462.

Oh, Y. M., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of Phonetics, 53*, 153–176. http://dx.doi.org/10.1016/j.wocn.2015.08.003.

Plauché, M., Delogu, C., & Ohala, J. J. (1997). Asymmetries in consonant confusion. In Proceedings of the 5th European conference on speech communication and technology (pp. 2187–2190).

Rabiner, L., & Juang, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall Inc..

Schatz, T. (2016). *ABX-Discriminability Measures and Applications*. École Normale Supérieure.

Schatz, T., Peddinti, V., Bach, F., Jansen, A., Hermansky, H., & Dupoux, E. (2013). Evaluating speech features with the Minimal-Pair ABX task: Analysis of the classical MFC/PLP pipeline. In Proceedings of interspeech. Retrieved from http://hal.archives-ouvertes.fr/hal-00918599/.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, 379–423.

Stevenson, S. (2015). *The strength of segmental contrasts: a study on Laurentian French*. University of Ottawa.

Stokes, S., & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language and Hearing Research, 48*, 577–592.

Strand, J., Simenstad, A., Cooperman, A., & Rowe, J. (2014). Grammatical context constrains lexical competition in spoken word recognition. *Memory & Cognition, 42*(4), 676–687. http://dx.doi.org/10.3758/s13421-013-0378-6.

Surendran, D., & Niyogi, P. (2003). Measuring the Usefulness (Functional Load) of Phonological Contrasts.

Surendran, D., & Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In O. Nedergaard Thomsen (Ed.), Competing models of linguistic change: Evolution and beyond (pp. 49–64).

van Severen, L., Gillis, J. J. M., Molemans, I., van den Berg, R., De Maeyer, S., & Gillis, S. (2013). The relation between order of acquisition, segmental frequency and function: the case of word-initial consonants in Dutch. *Journal of Child Language, 40*(4), 703–740. http://dx.doi.org/10.1017/S0305000912000219.

Versteegh, M. (2015). Spectral Retrieved from http://github.com/mwv/spectral.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *The Journal of the Acoustical Society of America, 54*(5), 1248–1266. http://dx.doi.org/10.1121/1.1914417.

Weber, A., & Smits, R. (2003). Consonant and vowel confusion patterns by American English listeners. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1427–1440).

Wedel, A. (2015). Biased variation shapes sound system change: Integrating data from modeling, experiments and corpora. Stuttgart, Germany.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech, 56*(3), 395–417. http://dx.doi.org/10.1177/0023830913489096.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition, 128*(2), 179–186. http://dx.doi.org/10.1016/j.cognition.2013.03.002.