CORTEX 151 (2022) 70–88

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special Issue "Mapping sound to meaning under challenging conditions": Research Report

The differential roles of lexical and sublexical processing during spoken-word recognition in clear and in noise

Antje Strauß [a], Tongyu Wu [b], James M. McQueen [b], Odette Scharenborg [c] and Florian Hintz [d,*]

[a] University of Konstanz, Konstanz, DE, Germany
[b] Radboud University, Nijmegen, NL, the Netherlands
[c] Multimedia Computing Group, Delft University of Technology, Delft, NL, the Netherlands
[d] Max Planck Institute for Psycholinguistics, Nijmegen, NL, the Netherlands

ARTICLE INFO

Article history:
Received 27 September 2021
Reviewed 28 November 2021
Revised 21 January 2022
Accepted 13 February 2022
Published online 16 March 2022

Keywords:
Spoken-word recognition
Background noise
Lexical decision
Lexical frequency
Neighborhood density
Phonotactic probability

ABSTRACT

Successful spoken-word recognition relies on interplay between lexical and sublexical processing. Previous research demonstrated that listeners readily shift between more lexically-biased and more sublexically-biased modes of processing in response to the situational context in which language comprehension takes place. Recognizing words in the presence of background noise reduces the perceptual evidence for the speech signal and – compared to the clear – results in greater uncertainty. It has been proposed that, when dealing with greater uncertainty, listeners rely more strongly on sublexical processing. The present study tested this proposal using behavioral and electroencephalography (EEG) measures. We reasoned that such an adjustment would be reflected in changes in the effects of variables predicting recognition performance with loci at lexical and sublexical levels, respectively. We presented native speakers of Dutch with words featuring substantial variability in (1) word frequency (locus at lexical level), (2) phonological neighborhood density (loci at lexical and sublexical levels) and (3) phonotactic probability (locus at sublexical level). Each participant heard each word in noise (presented at one of three signal-to-noise ratios) and in the clear and performed a two-stage lexical decision and transcription task while EEG was recorded. Using linear mixed-effects analyses, we observed behavioral evidence that listeners relied more strongly on sublexical processing when speech quality decreased. Mixed-effects modelling of the EEG signal in the clear condition showed that sublexical effects were reflected in early modulations of ERP components (e.g., within the first 300 msec post word onset). In noise, EEG effects occurred later and involved multiple regions activated in parallel. Taken together, we found evidence –

* *Corresponding author.*
E-mail address: florian.hintz@mpi.nl (F. Hintz).
https://doi.org/10.1016/j.cortex.2022.02.011

especially in the behavioral data – supporting previous accounts that the presence of background noise induces a stronger reliance on sublexical processing.

1. Introduction

Decades of research on spoken-word recognition have led to a number of different theoretical and computationally implemented models (e.g., Trace, McClelland & Elman, 1986; Shortlist A, Norris, 1994; Shortlist B, Norris & McQueen, 2008; PARSYN, Luce et al., 2000; EARSHOT, Magnuson et al., 2020; see Weber & Scharenborg, 2012, for review). While these models differ in important aspects, for example concerning the directionality of the flow of activation (Magnuson et al., 2018; McQueen et al., 2003), the majority of them agree that word recognition involves representations corresponding to (minimally) sublexical phonological structures, word forms and word meanings. There is also consensus that word recognition is cascaded, in the sense that information flows continuously through different stages of the recognition process, rather than consisting of serial stages where a later stage starts only after an earlier one has finished. The canonical view on spoken-word recognition therefore assumes a tight interplay between sublexical and lexical processing, such that – as speech is perceived and processed – activation flows from sublexical levels to lexical levels of processing. At the latter stage, multiple target candidates, consistent with the incoming phonological information, compete for recognition and as information accumulates, inconsistent target word candidates are excluded until the correct target is selected.

As with the majority of language comprehension studies, research on spoken-word recognition has been conducted predominantly in strictly controlled laboratory environments, using high quality audio recordings presented to the participants in shielded experimental booths. However, speech recognition outside the lab frequently takes place under adverse conditions (e.g., in crowded public places, in cars, trains or on airplanes) where the speech signal is distorted by the presence of background noise (Mattys et al. 2012, for review). Previous research on spoken-word recognition under adverse conditions showed that the presence of noise has profound effects on the underlying mechanisms and the neurobiological markers associated with these mechanisms. Building on this body of research, the present study was concerned with the interplay between sublexical and lexical processing, as there is agreement across most models that spoken-word recognition minimally involves processing at these stages. Specifically, we investigated whether listeners adjust their reliance on sublexical and lexical processing in the face of signal degradation and if/how the neurobiological markers associated with sublexical and lexical processing change accordingly.

1.1. Processing at sublexical and lexical levels

The fundamental process that underlies speech perception is translating continuous acoustic cues into abstract representational units that allow access to meaning. To achieve this feat, listeners must map the acoustic cues they perceive onto abstract, speech-specific sublexical representations (e.g., phonemes), which we summarize here as "sublexical processing". Furthermore, listeners must map these representations onto lexical representations, and onto semantic representations (e.g., Chen & Mirman, 2012; Hickok & Poeppel, 2007; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, & Cutler, 2000), which we summarize as "lexical processing".

Previous research has shown that the speed and accuracy with which spoken words are recognized are influenced by statistical regularities pertaining to the words' make-up. That is, earlier studies identified properties that listeners exploit when recognizing spoken words and that variation in these properties predict how quickly and accurately words are recognized. Three important properties are word frequency, neighborhood density and phonotactic probability. While word frequency is assumed to affect processing at the lexical level, phonotactic probability is assumed to affect processing at the sublexical level. Neighborhood density is considered a hybrid as it integrates statistical regularities pertaining to both levels.

1.2. Word frequency

Word frequency refers to the frequency of occurrence of whole word forms in a given language (see Brysbaert et al., 2018, for review). Effects of lexical frequency have been reported in numerous behavioral studies using for example speeded auditory lexical decision paradigms, where participants listen to spoken items and are instructed to indicate as fast as possible whether they heard a word of their language or not (Taft & Hambly, 1986; Luce & Pisoni, 1998; Cleland, Gaskell, Quinlan, & Tamminen, 2006, Ferrand et al., 2018). It is commonly found that words with higher frequency are recognized faster than words with lower frequency (e.g., Coltheart et al., 2001).

Furthermore, when presenting spoken words in the presence of background noise, research has shown that high-frequency words are recognized more accurately than low-frequency words (e.g., Broadbent, 1967; Howes, 1957; Morton, 1969; Pollack, Rubinstein, & Decker, 1960; Preston, 1935). The advantage in recognition accuracy and speed of high-frequency over low-frequency words was found to be robust in energetic masking (e.g., speech-shaped background noise), even across the lifespan (e.g., Van Engen et al., 2020). Moreover, a recent study found that frequency showed strong effects on listeners' misperceptions in energetic masking, but showed reduced or absent effects in informational masking (e.g., degradation of the target speech by irrelevant linguistic signals, Cooke et al., 2019).

Effects of word frequency have been reported in many neurocognitive studies as well using magneto- or electroencephalography (M/EEG). Frequency was found to modulate the N400, an M/EEG component peaking around 400 msec after word onset with centroparietal scalp distribution in EEG or a left temporal scalp distribution in MEG (for review see Lau et al., 2008; Kutas & Federmeier, 2011). Typically, words with lower frequency elicit enhanced N400 amplitudes compared to those with higher frequency. Variations in the N400 amplitude are considered to reflect variations in the processing efforts required to map phonological forms onto their lexical meanings (Van Petten & Kutas, 1990). Moreover, some previous studies observed 'late N400' effects during time regions after the spoken words' offsets (Desroches et al., 2009; Dufour et al., 2013). These frequency effects have been assumed to reflect the ease with which the target is selected as the best candidate, when the speech signal is no longer compatible with other lexical candidates (Desroches et al., 2009).

To our knowledge, no study has yet systematically investigated the neural markers of word frequency for spoken-word recognition in noise. Presumably, this is because most studies control for frequency rather than manipulate it. However, given the robustness of the behavioral advantage as reported above, it is rather likely that highly frequent words also reduce the N400 amplitude in adverse listening conditions.

1.3. *Phonological neighborhood density*

Phonological neighborhood density captures a word's phonological similarity to other existing words. It is typically operationalized by counting the number of words that can be formed from a target word by deletion, addition, and substitution of one phoneme (Landauer & Streeter, 1973). Like word frequency, phonological neighborhood density is canonically considered to have a lexical locus: Inhibitory effects for words with large similarity values are assumed to result from enhanced competition with similar sounding words. However, since the measure considers phonological (i.e., sublexical) properties of the target as well, its locus is controversial.

Previous experimental reports provided mixed results with neighborhood density having both inhibitory (lexical locus) and facilitatory (sublexical locus) effects. Thus, neighborhood density might be best conceived as a hybrid measure, with loci at lexical *and* sublexical levels. Specifically, while the majority of behavioral studies reported that high-density words were recognized less accurately and more slowly compared to low-density words (e.g., Goldinger, Luce, & Pisoni, 1989; Vitevitch & Luce, 1998; Vitevitch & Luce, 1999; Ziegler, Muneaux, & Grainger, 2003; see Vitevitch Luce, & 2016, for review), others like Vitevitch and Rodríguez (2005; see also Ferrand et al., 2018 for a study in French) have found that dense phonological neighborhoods produced a facilitatory rather than inhibitory effect in a lexical decision task in Spanish. Vitevitch and Rodríguez (2005) reasoned that differences in processing across English (prevailing language in most experiment) and Spanish might account for the reverse directionality of the neighborhood density effect. For example, it is conceivable that speakers of Spanish generally rely more strongly on sublexical than lexical processing. In doing so, neighborhood density effects might reverse, since words with dense phonological neighborhoods are often made up of high-frequency phonemes. On such an account, 'reverse neighborhood density effects' would be phonotactic probability effects in disguise.

Relatedly, when manipulating phonological neighborhood density in neurocognitive studies, contradictory results have been found concerning the modulation of the P200, a positive potential peaking approximately 200 msec after auditory stimulus onset. The P200 component has been associated with the processing of the stimulus' physical properties (Donchin, Ritter, & McCallum, 1978). Hence, when the P200 was found to be increased for high-density compared to low-density words, Dufour et al. (2013) speculated whether this was due to a confound (i.e., high-density words exhibiting higher phonotactic probability). In contrast, when the P200 has been shown to be decreased for high-density compared to low-density words, Winsler et al. (2018) associated neighborhood density effects with co-activation of larger sublexical and lexical networks leading to enhanced competition.

Despite the inconsistent results on the P200 component, it has been found robustly that words with more neighbors elicit larger N400 amplitudes than words with fewer neighbors. This has been argued to reflect enhanced efforts during lexical access due to the activation of many similar sounding words that compete with each other (Dufour et al., 2013; Winsler et al., 2018).

To our knowledge, only one study investigated the effects of neighborhood density on word recognition in the presence of background noise. As in clear speech conditions, Hunter (2016) found that high-density words elicit larger N400 amplitudes than low-density words.

1.4. *Phonotactic probability*

Phonotactic probability refers to the frequency with which a sublexical segment (e.g., a phoneme) or combinations of segments occur in a certain position within a word. As indicated above, phonotactic probability is sometimes positively correlated with phonological neighborhood density (Vitevitch, Luce, Pisoni, & Auer, 1999). As for word frequency, higher phonotactic probability during spoken-word recognition is associated with facilitatory effects—albeit at the sublexical level.

Advantages of high phonotactic probability have been reported in behavioral studies with tasks encouraging phonological processing, for example a same-different task when using non-words or phonologically neighboring words, where the increase of phonotactic probability led to faster and more accurate responses (Vitevitch & Luce, 1998; 1999, Exp. 1 & 2; Vitevitch et al., 1999; Hunter, 2013). However, when the task settings encouraged lexical processing, as for example in a lexical decision task, the effect of phonotactic probability disappeared (Dufour et al., 2013; Pylkkänen, Stringfellow, & Marantz, 2002; Vitevitch & Luce, 1999, Exp. 3).

In neurocognitive studies, the manipulation of phonotactic probability modulated the amplitude of the PMN, an early negative component peaking around 250 msec over frontocentral regions. The PMN is considered to reflect the

mismatch between expected and perceived phonological forms (Connolly & Phillips, 1994; for review and discussion of the PMN see; Lewendon et al., 2020). Dufour et al. (2013) found that words with high (compared to low) phonotactic probability elicited reduced PMN amplitudes suggesting facilitated sublexical processing. Moreover, it has been found that the PMN peaked earlier for high compared to low phonotactic probability words (Hunter, 2013; Exp. 1).

To our knowledge, the behavioral and neurocognitive impact of phonotactic probability has not yet been studied in the presence of background noise. Hence, it is unknown whether and how the manipulation of phonotactic probability would affect spoken word recognition in adverse listening conditions.

1.5. *Inducing modes of sublexical and lexical processing*

As discussed above, previous research showed that the nature of neighborhood density effects is variable, with neighborhood density showing inhibitory and facilitatory effects in certain situations. More globally, it appears that listeners can adjust their reliance on sublexical and lexical processing in response to situational and task demands. Further support for the notion that listeners can readily shift between modes of processing comes from a series of experiments by Vitevitch and Luce (1998, 1999). For example, Vitevitch and Luce (1998) presented English listeners with existing and non-existing English words, both varying in phonological neighborhood density, and asked participants to repeat the stimulus items as quickly as possible. The authors reasoned that listeners access and use lexical information when processing existing words but must rely on sublexical processing for non-existing words since lexical information is not available. The results revealed that dense-neighborhood words were repeated more slowly than sparse-neighborhood words, but that dense-neighborhood non-words were repeated more quickly than sparse neighborhood non-words. This behavior suggests that neighborhood density had inhibitory effects on words (reflecting stronger reliance on lexical processing) and facilitatory effects on non-words (reflecting stronger reliance on sublexical processing).

Vitevitch (2003) further showed that lexical and sublexical modes of processing are not confined to a specific type of target stimulus (lexical: words, sublexical: non-words) but can be induced by the situational context in which speech processing takes place. Specifically, the author presented existing English words varying in neighborhood density in experimental contexts where these were surrounded by fillers that were either predominantly existing or non-existing words. When presented with mostly existing words, participants relied on a lexical mode of processing, yielding inhibitory effects for words with dense neighborhoods. When the same words were presented in a context where the majority of words were non-existent, participants relied on a sublexical mode of processing, yielding facilitatory effects for words with dense neighborhoods.

More relevant to the present study is the previous work on speech recognition under adverse conditions showing that the presence of background influences the weighting of lexical and sublexical processing. In an effort to dissociate energetic from informational factors when recognizing speech under adverse conditions, Mattys, Brooks, and Cooke (2009; see also Mattys, White, & Melhorn, 2005; Newman, Sawusch, & Wunnenberg, 2011) provided evidence that the presence of background noise can also cause listeners to adjust their processing strategies. Their study focused on word segmentation and participants were presented with phrases whose combination between lexical structure and acoustic realization induced: (1) lexically driven segmentation (e.g., phrase perceived as 'mild option'), or (2) sublexically driven segmentation (phrase perceived as 'mile doption'), or (3) ambiguous segmentation [phrase perceived as lying in between (1) and (2)]. Participants' task was to judge whether they had heard 'mild' or 'mile' at the beginning of the phrase. Mattys and colleagues showed that participants were more likely to adopt a sublexically-driven segmentation strategy under energetic masking conditions (i.e., unintelligible background babble of noise). Importantly, the authors highlighted that based on the task they used, no claims about the effects of the relative time course of energetic (or informational) masking effects can be made since participants were under no pressure to make their judgments. Similarly, it is unclear whether the stronger reliance on sublexical processing in noise generalizes to speech comprehension tasks that (1) do not capitalize on segmentation and that (2) do not capitalize on local acoustic-phonetic contrasts (e.g., allophonic variation between 'mild option' and 'mile doption'). Finally, it is unclear how neural markers associated with lexical and sublexical processing change in the face of signal degradation.

1.6. *The present study*

The goal of the present study was to further investigate how the presence of background noise influences the interplay between lexical and sublexical processing during spoken-word recognition. Specifically, we studied whether listeners adjust their reliance on lexical- and sublexical-level processing in response to noise-induced uncertainty about the speech signal. To enable investigations of the time course and the neurobiological markers of the effects of noise, we conducted an EEG experiment. Our native Dutch participants carried out two experimental blocks. In the first block, they were presented with spoken Dutch words and non-words embedded in speech-shaped background noise. In the second block, we presented participants with the same words and non-words again but in the clear, which we used as a baseline condition. While the clear-speech baseline was the same for each participant, they heard the words in noise at one of three different SNRs. We used different SNRs to track the potential shifts in participants' processing strategies as the speech signal deterioration increased. SNR was implemented as a between-participants manipulation to maintain an acceptable balance between the number of trials needed for our ERP analyses and a workable number of trials for the participants. Note that we opted for speech-shaped background noise, a form of energetic masking, rather than informational masking for degrading our stimuli. The reason is that in the present study we were interested in perceptual uncertainty and its effects on spoken-word recognition (Mattys et al., 2012). Moreover, while there are data demonstrating how the effects

of word frequency change in energetic versus informational masking (e.g., Cooke et al., 2019), such systematic investigations are missing for phonological neighborhood density and phonotactic probability. Given the more extensive literature on these two predictors in energetic masking, we decided to use speech-shaped background to facilitate comparison with earlier studies.

Our participants were instructed to listen to each stimulus (one at a time) and carry out a two-stage task: Their first task was to decide as quickly as possible whether the presented stimulus was an existing Dutch word or not. In case of a yes-response, their second task was transcription (i.e., to type in the word they thought they heard). As has been argued elsewhere, lexical decisions may be carried out based on phonotactic patterns and/or phonotactic probabilities, *without* access to the lexicon (Balota & Chumbley, 1984, for discussion). To ensure that listeners access their lexicons, we added the transcription task, which required mapping the perceived acoustic signal onto an existing lexical representation. Non-word trials were excluded. For correct responses to words (correct lexical decision *and* correct transcription), we extracted lexical decision times. Moreover, these trials were submitted to EEG analyses.

We modelled each of the three dependent variables (accuracy, RT, ERP amplitude) using word frequency, phonological neighborhood density and phonotactic probability as predictors. In contrast to previous research, which mostly operationalized these variables in a dichotomous manner (high versus low), we used them as continuous predictors for better statistical power and to avoid arbitrary groupings of items.

1.7. *Predictions*

As outlined above and in line with most theories, we expected that listeners engage in sublexical and lexical processing during spoken-word recognition in clear and in noisy listening conditions: Incoming speech sounds are perceived and activate mental phoneme representations at sublexical levels, which send activation to lexical levels where words consistent with the incoming speech signal compete for recognition. As explained above, all three of our predictors (word frequency, phonological neighborhood density and phonotactic probability) are established predictors of spoken-word recognition and are assumed to have their locus at sublexical or lexical levels of representation (or at both). We reasoned that changes in the effects of these predictors across clear and noise conditions are most likely to reflect adjustments in listeners' reliance on sublexical and lexical processing, respectively.

In terms of simple effects, we hypothesized that we would replicate previous reports of lexical frequency effects across all listening conditions, with more accurate and faster responses for words of higher compared to lower frequency (e.g., Broadbent, 1967). Frequency effects are assumed to have their locus at the lexical level, with, according to one account, high-frequency word representations reaching critical activation levels faster/more accurately than low-frequency representations due to higher baseline activation (Coltheart et al., 2001).

Similarly, we expected to find the canonical inhibitory effects of phonological neighborhood density (predominantly localized at the lexical level) on accuracy and reaction times in the clear as well as in the noise conditions (lower accuracy and slower responses for words with dense rather than sparse neighborhoods, Luce & Pisoni, 1998). Concerning effects of neighborhood density in noise, a distorted speech signal is likely to reduce the resolution of individual phonemes, thereby increasing the set of potential target candidates. In noise, words residing in dense phonological neighborhoods are thus more likely to be confused with similar sounding competitors than words residing in sparse neighborhoods. It is conceivable that such effects on accuracy and RTs are moderated by SNR such that as the speech quality decreases, accuracy decreases and RTs increase as a function of a widening competitor space.

It was unclear whether we would observe phonotactic probability effects in the clear given that previous reports of such effects were confined to specific tasks (e.g., same-different task). However, in line with previous proposals (Mattys et al., 2009), we predicted simple effects of phonotactic probability in noise. Under noisy conditions, listeners are assumed to rely more strongly on sublexical processing and words with higher compared to lower phonotactic probability should be recognized more accurately and faster. Specifically, with noise reducing the resolution of individual phonemes, we expected listeners to experience facilitation in recognizing words in noise that are composed of frequent rather than infrequent phoneme combinations since listeners may exploit statistics about the frequency of occurrence of a given phoneme in a given position in case of noise-induced ambiguity. We predicted the effects of phonotactic probability to become stronger as the speech quality became worse (SNR decreased). In a similar vein, one may predict *facilitatory* effects of neighborhood density on accuracy and RTs in noise. That is, given the positive correlation between neighborhood density and phonotactic probability (i.e., words residing in dense phonological neighborhoods are often composed of frequent phoneme combinations), it is conceivable that with noise reducing individual phoneme resolution listeners weigh sublexical more strongly than lexical processing such that the facilitatory effects of dense-neighborhood words' high phonotactic probability at the sublexical level trump the inhibitory effects of enhanced competition at the lexical level.

One unique feature of the present study concerned investigating the interactions between the three predictors. One study that laid important ground work in that regard (pertaining to frequency and neighborhood density), was conducted by Benkí (2003). Benkí (2003) presented US American participants with English CVC words and non-words in noise at four different SNRs. Both words and non-words were phonetically balanced such that the ten phonemes in the initial position used in the stimulus set, the ten vowels and the ten phonemes in the final position were evenly distributed. Participants carried out a transcription task. In addition to (phoneme) recognition accuracy, Benkí (2003) focused on j-factor models for the analysis of participants' responses. j-factor models quantify the relationship between the recognition of a whole (word/non-word) and the recognition of its parts. Benkí found that participants perceived the non-words as having three independent phonemes (j-factor of 3); words were found to be perceived as having 2.34 independent units.

Among the words, high-frequency words were perceived as having fewer independent units than low-frequency words. That is, high-frequency words were perceived 'more holistically' than low-frequency words. Moreover, words with dense phonological neighborhoods were perceived as having .5 more independent units than words with sparse neighborhoods (i.e., sparse-neighborhood words were overall perceived 'more holistically' than dense-neighborhood words).

On this account, one may conjecture that CVC-words that are generally perceived as consisting of more independent phonemic units (i.e., low-frequency and dense-neighborhood words) are more susceptible to be moderated by variations in phonotactic probability than words that are generally processed more holistically. This might be especially so under sub-optimal listening conditions. Against this background, we additionally predicted interactions between phonotactic probability and frequency (higher accuracy and faster RTs for low-frequency words composed of common rather than uncommon phoneme sequences) and between phonotactic probability and neighborhood density (higher accuracy and faster RTs for dense-neighborhood words composed of common rather than uncommon phoneme sequences). The latter prediction resonates with the notion that neighborhood density can have both inhibitory and facilitatory effects in certain situations (cf. work by Vitevitch and collaborators).

In terms of EEG correlates, we predicted reduced N400 amplitudes elicited by high-compared to low-frequency words. Moreover, in line with the canonical view, we expected words with many compared to few neighbors to elicit larger N400 amplitudes in clear and noisy listening conditions (due to the enhanced competition among similar sounding words). Neighborhood density effects might also be observed on early ERP effects, namely the P200 component. However, the directionality of such effects was uncertain (cf. Dufour et al., 2013; Winsler et al., 2018) and whether such effects would emerge in the presence of background noise. With regard to EEG effects, phonotactic probability previously showed effects on the PMN, we therefore predicted reduced amplitudes for high-compared to low-probability words (reflecting facilitation). These effects might increase with decreasing speech quality.

2. Methods

The spoken stimuli used in the experiment, the raw data and analysis code can be accessed from the Max Planck Institute for Psycholinguistics Archive: https://hdl.handle.net/1839/1fbdb007-96c9-40cb-8031-46af2e019fac. The study design was not pre-registered. We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. No part of the study analyses was pre-registered prior to the research being conducted, nor were inclusion/exclusion criteria.

2.1. Participants

Thirty-one members of the subject pool of Radboud University (all students or recent alumni, twelve males, mean age = 23, *SD* = 3, range = 18–31) took part in the experiment. All were native speakers of Dutch, right-handed, and did not report any history of learning or language disabilities or neurological or psychiatric disorders. The participants were given a €20-voucher as compensation for their participation. The ethics board of the Faculty of Arts at Radboud University approved the study. One participant had to be excluded from the analysis due to an experimental error; the final set comprised 30 participants. The sample size was determined on the basis of similar previous studies (e.g., Dufour et al., 2013).

2.2. Materials

We selected 160 monosyllabic Dutch words from the Subtlex-NL database (Keuleers et al., 2010). All of them had a CVC-structure and their recognition and uniqueness points after the third phoneme. The words varied substantially in word frequency, that is their frequency of occurrence per one million words (operationalized as Zipfian frequency,[1] *ZipfF*; Van den Heuvel et al., 2014; *M* = 4.18; *SD* = 1.05; range = 2.15–7.34), and in phonological neighborhood density (*ND*; *M* = 22.23; *SD* = 8.68; range = 4–53; measured using Clearpond; Marian et al., 2012). *ND* was defined as the number of words that can be formed from the target word by adding, deleting, or substituting a single phoneme (Landauer & Streeter, 1973; Luce & Pisoni, 1998). As stressed by Luce and Pisoni (1998; see also Newman et al., 1997), it is important to incorporate the lexical frequency of a target's neighbors. After having determined the number of neighbors for each item, we therefore followed Newman et al. (1997) and weighted the neighbors by their log-transformed frequencies and summed them to yield a frequency-weighted neighborhood density (*FWND*; *M* = 24.64, *SD* = 11.85, range = .31–69.08). Finally, the 160 CVC target words varied in phonotactic probability (Vitevitch & Luce, 2004), which we operationalized as triphone frequency (*TriF*, i.e., the sum of the positional single-phoneme probabilities as retrieved from Clearpond, Marian et al., 2012; *M* = .16, *SD* = .04; range = .04–.31). Table 1 provides an overview of the descriptive statistics of the three variables as well as the correlations between them.

Based on this set of 160 Dutch words, we constructed 160 non-words using the non-word generator Wuggy (Keuleers & Brysbaert, 2010). Specifically, one segment in each original word was changed without violating phonotactic constraints of Dutch to yield a non-word (see Table 1). Word and non-word stimuli were spoken by a female native speaker of Dutch at a normal pace with neutral intonation in a sound-shielded booth. Recordings were made using a Sennheiser microphone sampling at a frequency of 44.1 kHz (16-bit resolution). The individual words and non-words were cut using Audacity® version 2.05 (Audacity Team, 2014). The recordings of the spoken words were on average 447 msec long (*SD* = 78 msec, range = 270–750 msec); the recordings of the non-words were on average 451 msec long (*SD* = 72 msec, range = 294–636 msec), as measured using Praat (Boersma, 2001).

We created three additional versions of each sound file using Praat (Boersma, 2001) by adding speech-shaped background noise at three different SNRs to each word and non-

[1] Calculated as *log10(frequency per million words) + 3*.

Table 1 – Parameter descriptions and correlations.

Variable	Original (n = 160)	Excluded (n = 23)	Included (n = 137)	Non-words (n = 160)	Correlations (n = 137)	
	M (SD) Range	*M (SD)* Range	*M (SD)* Range	*M (SD)* Range	1. *ZipfF*	2. *FWND*
1. *ZipfF*	4.18 (1.05)	3.38 (.78)	4.32 (1.03)	–		
	2.15-7.34	2.20-4.73	2.15-7.34	–		
2. *FWND*	24.64 (11.85)	22.31 (9.85)	25.03 (12.14)	17.42 (9.64)	.23	
	.31–69.08	8.82-48.45	.31–69.08	.37–44.40	[.07, .38]	
3. *TriF*	.16 (.04)	.18 (.04)	.15 (.04)	.17 (.06)	.18	.38
	.04-.31	.11-.26	.04-.31	.07-.32	[.01, .34]	[.23, .52]

Note. *ZipfF* for Zipfian frequency, *FWND* for frequency-weighted neighborhood density, *TriF* for triphone frequency; *M* and *SD* abbreviate mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for correlations (Cumming, 2014). Descriptive statistics are reported for the full material set (n = 160) and for the items used in the analysis (n = 137). Twenty-three items were excluded based on low recognition accuracy (see Item analysis section; cf. Dufour et al., 2013).

word. To that end, recordings were down-sampled to 16 kHz to match the sampling frequency of the to-be-added noise. The noise was ramped up over a period of 205 msec before the onset of each lexical item and ramped down over 205 msec after the offset of each lexical item. The ramping up was included for participants to get accustomed to the presence of noise before being presented with the relevant speech. Similarly, ramping down ensured a more pleasant and natural ending of the stimulus (i.e., in many cases of constant background noise in the real world, speech may stop while noise continues).

Speech-shaped noise was added at three different SNRs: + 10dB, + 6dB, + 2dB. When selecting the noise intensity level for the between-participants manipulation, we aimed for a pattern where the easiest noise condition (highest SNR value) would yield a substantial decrease in word recognition accuracy compared to the clear condition and where the most difficult SNR condition (lowest SNR value) would yield sufficient numbers of correctly recognized trials for our EEG analyses. Moreover, performance should differ substantially between SNRs. An earlier behavioral study conducted in our lab, involving 44 native speakers of Dutch, using the same materials and masker as in the present study, suggested a linear decrease in transcription accuracy (clear: 96%, SNR + 6dB: 75%, SNR + 2dB: 61%, SNR-2dB: 39%, SNR-6dB: 23%, Hintz et al., 2016; see also Scharenborg et al., 2018). We deemed 61% (i.e., 98 of 160 trials per participant, SNR + 2dB) sufficient as the lower limit of the noise intensity manipulation in the present study, leaving enough data points for the ERP analysis (Dufour et al., 2013). Moreover, we deemed differences in recognition accuracy ranging between 14% and 20% between SNR conditions in our earlier study sufficient to detect substantial differences between SNR conditions in the present study. We therefore decided to operationalize SNR conditions in the present study to decrease in increments of four, starting at + 10dB and ending at + 2dB. The original, noise-free sound files were used in the clear condition and served as baseline.

2.3. *Procedure*

Participants were tested individually in a sound-shielded booth. They were seated in a relaxed position in front of a 19-inch CRT screen. The experiment was implemented in Presentation®. Participants carried out six practice trials (three words and three non-words) before the start of the experiment. For the first experimental block, participants were randomly assigned to one of the three SNR conditions (+ 2dB, + 6dB, or + 10dB, ten per SNR). They listened to the complete set of 320 items (160 words and 160 non-words) at a fixed SNR. For the second experimental block, participants were presented with the same 320 stimuli as in the first block but without background noise (i.e., clear condition). The order of items within noise and clear blocks was randomized for each participant. Breaks could be taken every 80 trials (six breaks in total). The entire session, including EEG preparation, took about 2 h.

On each trial, a fixation cross was presented in the middle of the screen for 300 msec followed by the playback of the audio file. Participants were instructed to carry out a lexical decision task ('Is the spoken stimulus an existing Dutch word?'). Recall that all existing words had their recognition and uniqueness points after the third phoneme. Thus, participants had to listen to each word in its entirety to be able to make a decision. The left-right assignment of yes/no response buttons was counterbalanced across participants. In case of a yes-response, participants were asked to type-in the word they heard. This was done to ensure that participants recognized the correct word in question. Obvious typos were corrected (e.g., misspellings where adjacent keys were accidently pressed resulting in a non-word: 'bae' instead of 'bar'). Accuracy was operationalized as the proportion of correct responses (yes-decision followed by a correct transcription). Reaction times were recorded from the offset of each lexical stimulus.

2.4. *EEG acquisition*

EEG was recorded continuously from 59 active Ag/AgCl electrodes mounted in a cap according to the 10–20 system (Klem, Lüders, Jasper, & Elger, 1999). The signal was amplified using a Biosemi active amplifier with a bandpass filter of .016–100 Hz, sampling at a frequency of 1000 Hz, online referenced to the left mastoid. To monitor participants' eye movements, electrooculogram was recorded bipolar at a horizontal (left and

right eye corner) and a vertical (above and below left eye) line. A fifth additional electrode, placed on the right mastoid, was used as an extra reference, used offline during the analysis. All electrode impedances were kept below 20 kΩ.

2.5. Initial item analysis

Prior to pre-processing the EEG data and prior to any inferential statistics, we conducted a descriptive analysis on the word items to assess their overall accuracy (i.e., over all participants, based on lexical decision *and* transcription). In case an item had an overall accuracy below 60% in the clear condition (Dufour et al., 2013), we removed that item from further analysis. This affected 23 of the 160 items (1380 trials in total). The remaining dataset thus contained 8220 trials (30 participants * 137 items * 2 listening conditions). As can be seen in Table 1, item exclusion mostly concerned low-frequency words (cf. Hintz et al., 2016).

2.6. Behavioral analysis: accuracy

Based on the responses to the remaining 137 items, we calculated recognition accuracy for the four listening conditions (clear, SNRs + 10dB, + 6dB, + 2dB, Table 2). In R (R Core Team, 2012), mixed-effects logistic regression models were fitted using the logistic linking function with accuracy as a binary dependent variable (1 = correct; 0 = incorrect). Continuous predictors, namely *ZipfF*, *FWND* and *TriF*, were centered and scaled to minimize multicollinearity. Continuous predictors and their interactions were included in the model. For the categorical predictor (fixed factor) *listening condition*, we adopted Helmert contrasts. Helmert contrast-coding compares the mean of one level of a fixed factor to the mean of the subsequent levels of the variable. For the present analysis, the first Helmert contrast [*SNR (H.1)*] compared the mean accuracy of SNR + 6dB against SNR + 10dB. The second Helmert contrast [*SNR (H.2)*] compared the mean accuracy of SNR + 2dB to the mean of SNR + 6dB and SNR + 10dB. The third Helmert contrast [*SNR (H.3)*] compared the mean accuracy of all clear trials to the mean of all noise trials, i.e., including SNR + 10dB, SNR + 6dB and SNR + 2dB. While the former two contrasts enabled us to capture the linear decrease in recognition accuracy as a function of decreasing SNR, the latter contrast facilitated the global comparison of clear and noise conditions. The model further contained random factors for participants and words (both with random intercepts). Including random slopes for *listening condition* (i.e., categorical variable contrasting noise with clear) resulted in a singular fit. Using the anova()-function, we selected the next, more parsimonious model that did not result in a singular fit (achieved through dropping the by-participants random slope).

Table 2 – Behavioral word recognition performance.

Listening condition	Accuracy			RT		
	N	M	SD	N	M	SD
Clear	4110	.90	.29	3572	494	225
SNR + 10 dB	1370	.81	.40	1057	622	324
SNR + 6 dB	1370	.73	.45	970	724	340
SNR + 2 dB	1370	.60	.49	787	716	379

Note. RTs in milliseconds. RT means and SDs based on trials included in the re-fitted linear mixed-effects models (extreme values excluded).

2.7. Behavioral data analysis: reaction times

RTs for correct responses to word items were modelled for clear and noise conditions separately. Visual inspection had suggested that the overall distribution of RTs was quite different in both conditions, most likely due to the presence of leading and trailing noise in the noise condition, motivating separate models for all clear and all noise trials. Prior to fitting the models, we removed trials with (log-transformed) RTs that were 2.5 standard deviations away from a participant's mean (in total: n = 116; 1.75%). As for the model of the accuracy-data, each RT-model contained the continuous predictors *ZipfF*, *FWND*, *TriF* (all scaled and centered), and their interactions. The RT_{noise}-model additionally contained *SNR* as a fixed factor. As before, we used Helmert contrast-coding for comparing performance across the three *SNR* levels. The first Helmert contrast [*SNR (H.1)*] compared the mean reaction times of SNR + 6dB against SNR + 10dB. The second Helmert contrast [*SNR (H.2)*] compared the mean reaction times of SNR + 2dB to the mean of SNR + 6dB and SNR + 10dB. As before, random effects for participants and words (both with random intercepts were included). Adding a random slope for *SNR* by-item resulted in a singular fit; the random slope was dropped.

After the first model fit, data points that were further away than 2.5 standard deviations from the model's fitted values (Baayen & Milin, 2010), were classified as outliers (noise model: 1.54% of all trials; clear model: 2.14% of all trials). These outliers were removed, and the model was refitted.

2.8. EEG data pre-processing

EEG data pre-processing for trials with correct responses to words was done in MATLABTM using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Owing to a technical error (trigger codes were not sent due to buffer overflow in the EEG recording system), three trials had to be excluded. In total, 6621 trials went into pre-processing (clear condition: 3716 trials, SNR + 10dB: 1097 trials, SNR + 6dB: 993 trials, SNR + 2dB: 815 trials). To avoid filter edge artifacts in the time window of interest, long epochs (±3 sec around word onset) were used for re-referencing to the average of both mastoids and for applying a bandpass filter of .1–50 Hz. Then, shorter epochs (including the baseline and the time window of interest from −1 sec to 2 sec around word onset) were selected. Next, a four-step artifact rejection process was applied.

First, an initial visual inspection was done to mark broken channels producing artifacts over many trials. Second, by using independent component analysis eye blinks and lateral eye movements were removed from the signal. In a third step, muscle artifacts were automatically identified via Fieldtrip routine and respective trials excluded when exceeding a threshold of ±200 μV. In the fourth and last step, data were visually inspected again to detect possible remaining artifacts.

Finally, the channels that had been marked as broken were interpolated (concerning .93% of all trials). In total, 95 trials (1.43% of the ingoing data) were excluded during pre-processing.

2.9. EEG data analysis: cluster-based permutation

We started our EEG data analyses by using a data-driven approach, namely cluster-based permutation testing as implemented in Fieldtrip (Oostenveld et al., 2011). This approach was necessary given the sparse literature on EEG effects of ZipfF, FWND and TriF during spoken-word recognition in clear and in noise. In doing so, we also avoided subjective biases when selecting regions and times of interest for our subsequent mixed-model analyses. As for RTs, we analyzed clear and noise conditions separately, as the time course of grand-average ERPs were substantially different for the two listening conditions (i.e., as acoustic stimulation, through the presence of leading noise, began about 200 msec earlier in the noise condition, brain responses were shifted). Furthermore, as per our hypotheses, we predicted that the three properties (ZipfF, FWND and TriF) would play different roles in clear and in noise environments (e.g., showing effects at different points in time).

Single-trial time-domain EEG data were submitted to a multi-level or 'random effects' statistics approach (see Strauß et al., 2014). On the first level, i.e., for each individual separately, massed independent samples regression coefficient *t*-tests with contrast weights as independent variable (either ZipfF, FWND or TriF) were calculated. Uncorrected regression t-values and betas were obtained for all time–channel bins. On the second or group level, t-values were tested against zero in a two-tailed dependent samples *t*-test. A Monte-Carlo non-parametrical permutation method (1000 randomizations), as implemented in Fieldtrip, estimated type I-error controlled cluster significance probabilities ($\alpha = .025$). In total, we ran six separate tests to search for time-channel clusters: 3 predictors (ZipfF, FWND, TriF) x 2 listening conditions (clear, noise). Based on these six tests, we chose regions and times of interest (ROIs and TOIs), over which we averaged the EEG data for subsequent modelling in R.

2.10. EEG data analysis: mixed-effects modeling

The cluster-based permutation testing was followed up by further statistical evaluation using linear mixed-effects modeling. To that end, informed by the results of the cluster-based permutation analysis, we extracted the raw (i.e., not baseline corrected) single-trial data of regions and time windows of interest and averaged them across the respective ROI and TOI. Furthermore, we extracted a baseline (−200 msec–0 msec before word onset) for each trial in the respective ROI to be included as a covariate in the linear mixed-effects models (Alday, 2019). We used the lme4 package (v.1.1–23, Bates et al., 2015) in R to fit the various models. Below we list the two model templates used for analyses of the (1) noise and (2) clear data. As in the previous models, statistical properties, included as continuous predictors, were scaled and centered.

(1) *lmer [EEG ~ BaselineEEG * SNR * WordProperty + (1 | Participant) + (1 + SNR | Word), Data=eeg.noise]*
(2) *lmer (EEG ~ BaselineEEG * WordProperty + (1 | Participant) + (1 | Word), Data=eeg.clear]*

Both types of models contained random effects (both with random intercepts) for participants and words. The models for the noise data, additionally contained *SNR* as a categorical predictor, Helmert contrast-coded as for the RT analysis. The noise models further contained random slopes for *SNR*.

3. Results

3.1. Behavioral performance: accuracy

Average recognition accuracy for the 137 items for the different listening conditions is summarized in Table 2. As is to be expected, recognition accuracy was highest for the clear condition. Moreover, the means suggest a linear decrease in recognition accuracy as a function of decreasing *SNR*. To assess variability between participants and to screen for outliers, we also calculated mean recognition accuracy for each participant. None of the participants scored below 60% in the clear condition, indicating task compliance.

Table 3 summarizes the results of the mixed-effects model on word recognition accuracy, which we will discuss following the order of effects as listed in the table (from top to bottom). First, in line with numerous previous studies, word recognition accuracy was found to be positively affected by lexical frequency (*ZipfF*), such that words with higher lexical frequency were more often identified correctly than words with lower frequency.

Second, *listening condition* explained significant amounts of variance in the accuracy data (all three contrasts, *SNR [H.1]*, *SNR [H.2]*, and *SNR [H.3]*, were significant. *SNR [H.1]* compared the mean accuracy of *SNR* + 6 dB against *SNR* + 10 dB. *SNR [H.2]* compared the mean accuracy of *SNR* + 2 dB to *SNR* + 6 dB and *SNR* + 10 dB. *SNR [H.3]* compared the mean accuracy of clear trials against all noise trials (including *SNR* + 10 dB, *SNR* + 6 dB and *SNR* + 2 dB). *SNR [H.1]* and *SNR [H.2]* contrasts thus demonstrated that word recognition accuracy decreased as *SNR* decreased. Moreover, the *SNR [H.3]* contrast showed that word recognition was overall more accurate in clear than in noisy listening conditions.

Third, we observed four significant interactions. The first was between triphone frequency and Zipfian frequency (*TriF* × *ZipfF*). In order to understand its nature, we visualized it and the two significant three-way interactions that these factors were also part of, TriF x ZipfF x Listening Conditions and the TriF x ZipfF x FNWD. See Fig. 1. Fig. 1A plots accuracy as a function of triphone frequency (*TriF*) split into low, medium and high lexical frequency (*ZipfF*).[2] We observed the steepest slope (or highest gain) of increasing triphone frequency when lexical frequency was low. This finding can be further differentiated by considering the three-way interaction with neighborhood density (*TriF* × *ZipfF* × *FWND*, Fig. 1B)

[2] Note that the three-way split was adjusted for visualization purposes only.

Table 3 – Summary of model results on word recognition accuracy (clear and noise conditions).

	Estimate	SE	z	p
(Intercept)	1.66	.14	11.56	.000***
ZipfF	.47	.13	3.59	.000***
FWND	.17	.13	1.34	.179
TriF	.24	.14	1.70	.090
SNR [H.1]	−.28	.10	−2.93	.003**
SNR [H.2]	−.43	.06	−7.61	.000***
SNR [H.3]	.34	.03	9.93	.000***
TriF × ZipfF	−.37	.16	−2.36	.018*
TriF × FWND	.21	.12	1.72	.085
ZipfF × FWND	.14	.13	1.07	.283
TriF × SNR[H.1]	.10	.09	1.13	.260
TriF × SNR[H.2]	.02	.05	.44	.662
TriF × SNR[H.3]	−.07	.04	−1.79	.073
ZipfF × SNR[H.1]	−.05	.08	−.64	.520
ZipfF × SNR[H.2]	.07	.05	1.47	.142
ZipfF × SNR[H.3]	.00	.03	−.13	.893
FWND × SNR[H.1]	.07	.08	.96	.335
FWND × SNR[H.2]	.08	.05	1.66	.096
FWND × SNR[H.3]	−.02	.03	−.58	.559
TriF × ZipfF × FWND	−.42	.13	−3.17	.002**
TriF × ZipfF × SNR[H.1]	−.03	.09	−.36	.716
TriF × ZipfF × SNR[H.2]	−.08	.06	−1.50	.133
TriF × ZipfF × SNR[H.3]	.09	.04	2.18	.030*
TriF × FWND × SNR[H.1]	.11	.08	1.39	.164
TriF × FWND × SNR[H.2]	.05	.05	1.01	.312
TriF × FWND × SNR[H.3]	.04	.04	1.03	.302
ZipfF × FWND × SNR[H.1]	.04	.08	.54	.590
ZipfF × FWND × SNR[H.2]	.02	.05	.39	.699
ZipfF × FWND × SNR[H.3]	−.07	.04	−1.92	.055
TriF × ZipfF × FWND × SNR[H.1]	−.03	.08	−.39	.694
TriF × ZipfF × FWND × SNR[H.2]	−.03	.05	−.58	.565
TriF × ZipfF × FWND × SNR[H.3]	.11	.04	3.19	.001**

Note. Model formula: glmer [*Accuracy ~ ZipfF * FWND * TriF * SNR* + (1 | *Participant*) + (1 + *Noise versus Clear* | *Word*), family = binomial, data = df, glmerControl(optimizer = c("bobyqa"))]. *SNR [H.1]* compared the mean accuracy of *SNR* + 6dB against *SNR* + 10dB, *SNR [H.2]* compared the mean accuracy of *SNR* + 2dB to *SNR* + 6dB and *SNR* + 10dB, *SNR [H.3]* compared the mean accuracy of all clear trials against all noise trials, i.e., *SNR* + 10dB, *SNR* + 6dB and *SNR* + 2dB. *** denotes $p < .001$, ** denotes $p < .01$, * denotes $p < .05$.

demonstrating that effects of *TriF* and *ZipfF* were modulated by *FWND*: When lexical frequency and triphone frequency were low, high neighborhood density exerted a strong inhibitory effect on recognition accuracy (left-most plot in Fig. 1B). Interestingly, Zipfian frequency had strong effects on accuracy when words had low or mid-range triphone frequency (see brownish line in left-most and middle plot of Fig. 1B). When triphone frequency was high (right-most plot in Fig. 1B), accuracy was highest for low-frequency words with many phonological neighbors.

The *TriF*- and *ZipfF*-effects summarized above were independent of *listening condition*. However, both properties also interacted with *SNR* (*TriF* × *ZipfF* × *SNR[H.3]*[3]). As can be seen in Fig. 1C, the general pattern was that the poorer the signal-to-noise ratio, the more listeners benefited from triphone frequency. In clear speech, triphone frequency did not show effects, presumably due to close-to-ceiling performance and a strong reliance on lexical processing. As listening became harder, triphone frequency gained in importance, especially for words with low Zipfian frequency (compare green lines from left to right plots in Fig. 1C).

3.2. *Behavioral performance: reaction times*

We further examined the effects of the three statistical properties on lexical decision times. We only considered correct responses and – because of their substantially distinct distribution – modeled RTs of clear and noise conditions separately. Descriptive statistics can be found in Table 2. As can be seen, RTs (measured from word offset) were shortest for the clear condition (494 msec). When words were masked with noise at an SNR of + 10dB, RTs were on average more than 100 msec longer (622 msec). The RTs for both SNRs + 6 dB (724 msec) and + 2 dB (716 msec) were on average 100 msec longer than that of SNR + 10. The results of the mixed-effects models are summarized in Fig. 2 and Table 4.

For clear speech, we found a main effect of Zipfian frequency: the higher word frequency, the lower RTs. Moreover, we found a two-way interaction between *ZipfF* and *FWND* (*FWND* × *ZipfF*), as well as a three-way interaction between *ZipfF*, *FWND* and *TriF* (*TriF* × *FWND* × *ZipfF*). Both interactions are visualized in Fig. 2. As can be seen in Fig. 2A, reaction times increased for high-frequency words as *FWND* increased suggesting an increase in lexical competition. The reverse was found for low-frequency words: RTs decreased with increasing *FWND*.

This interaction was further differentiated by the three-way interaction as depicted in Fig. 2B. The plots highlight the two natures of *FWND*, which are most clearly visible in the high-*TriF* plot, where higher *FWND* had an inhibitory effect for high-frequency words and a facilitatory effect for low-frequency words. For mid-range *TriF* words, this cross-over pattern was pronounced less clearly and it was absent for low-*TriF* words. Instead, words with low triphone frequency generally displayed inhibitory effects with increasing neighborhood density.

As for the clear trials, we observed a main effect of Zipfian frequency when words were presented in noise, with faster RTs for words with higher frequency. Next to this main effect, we observed a two-way interaction involving Zipfian frequency and *SNR* (*ZipfF* × *SNR*; see Fig. 3C): At an intermediate level of signal degradation, i.e., at +6dB, there was less benefit of lexical frequency for RTs, relative to +10dB or at +2dB. Moreover, the analysis revealed a three-way interaction between Zipfian frequency, *SNR*, and triphone frequency (*TriF* × *ZipfF* × *SNR*). As illustrated in Fig. 3D, when words were recognized in more difficult listening conditions (i.e., +2dB), recognition speed for low-frequency words profited from high triphone frequency, suggesting a strong reliance on sublexical features (i.e., triphone frequency). In contrast, when recognizing low-frequency words at a higher, less difficult *SNR* (i.e., +10 dB) RTs showed the reverse pattern (RTs did not benefit from high triphone frequency and even increased). At an

[3] Note that we limited our visualization and interpretation to three-way interactions.

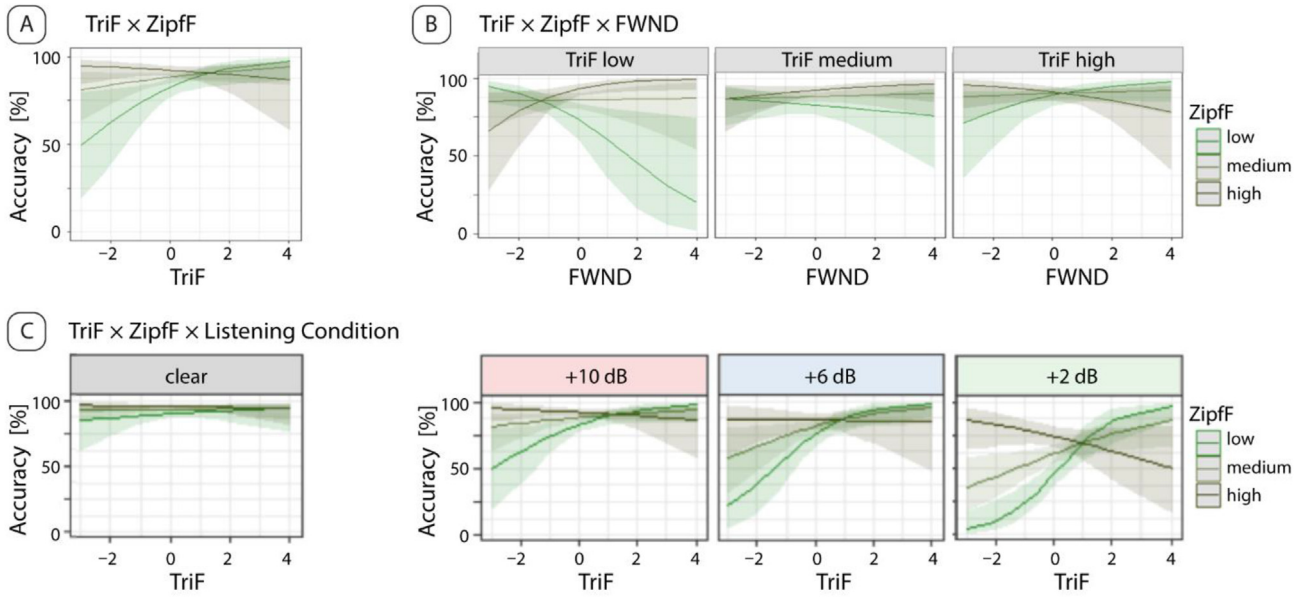

Fig. 1 – Mixed-effects model results for word recognition accuracy. For illustration purposes only, the continuous variables ZipfF and TriF were split into low (M < −1 SD), medium (M ±1 SD), and high (M > + 1 SD). Shaded areas represent the 95% confidence interval. A. Illustration of the interaction of phonotactic probability (TriF) and lexical frequency (ZipfF). In general, word recognition accuracy dropped when both measures were low. B. Illustration of the three-way interaction of TriF, ZipfF and frequency-weighted neighborhood density (FWND): high FWND had inhibitory effects (i.e., led to lower word recognition accuracy), when phonotactic probability (TriF) and lexical frequency (ZipfF) were low. C. Illustration of the three-way interaction of phonotactic probability (TriF), lexical frequency (ZipfF), and Listening Condition (SNR): The lower the SNR, the higher the gain of TriF in infrequent words.

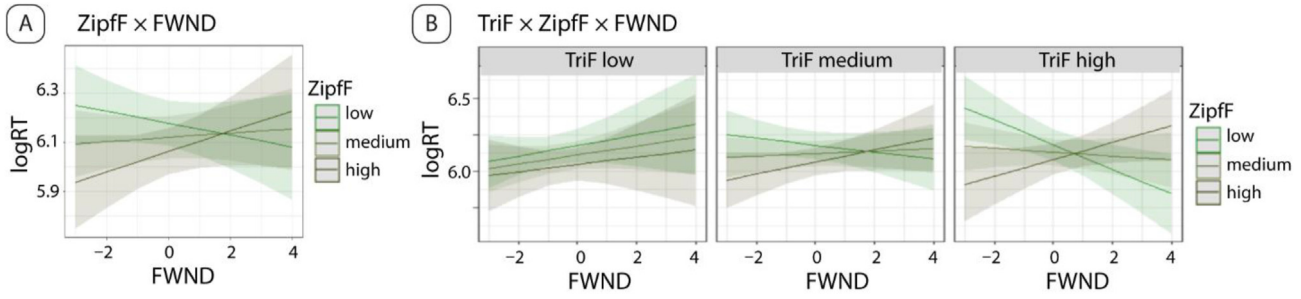

Fig. 2 – Model results for reaction times in clear speech. For illustration purposes only, the continuous variables ZipfF and TriF were split into low (M < −1 SD), medium (M ±1 SD), and high (M > +1 SD). Shaded areas represent the 95% confidence interval. A. Illustration of the interaction of lexical frequency (ZipfF) and frequency-weighted neighborhood density (FWND). In general, higher FWND had an inhibitory effect, i.e., it led to slowed reaction times, when ZipfF was high. On the other hand, higher FWND had a facilitatory effect, when ZipfF was low. B. Illustration of the three-way interaction of phonotactic probability (TriF), ZipfF, and FWND: When TriF was high, the interaction of FWND and ZipfF was strongest.

intermediate level of signal quality (i.e., +6dB), both lexical frequency and triphone frequency did not appear to interact.

Furthermore, we observed a two-way interaction of *TriF* and *FWND*. While high-*TriF* words appeared not to be modulated by *FWND*, RTs for mid-range and low-*TriF* words increased as *FWND* increased. Interestingly and comparable to the accuracy analysis (and to the RTs in clear speech), Zipfian frequency was found to interact with *TriF* and *FWND* in a three-way interaction (*TriF* × *FWND* × *ZipfF*). As depicted in Fig. 3B (note the similarity to Fig. 2B), the higher the triphone frequency, the stronger the benefit of neighborhood density, i.e., faster reaction times, for low lexical frequency items.

In sum, the behavioral data demonstrated that lexical frequency exerted a strong influence on both recognition accuracy and speed, both in the clear and in the presence of background noise. All models revealed main effects of *ZipfF* with higher accuracy and faster RTs for words of higher frequency. With regard to low-frequency words, we observed interesting patterns suggesting that sublexical features were crucial for determining recognition accuracy and speed. That is, we found that low-frequency words, which exhibited high triphone frequency benefitted from a high neighborhood density (in terms of accuracy, Fig. 1B right-hand plot). Second, under ideal listening conditions, low-frequency words were

Table 4 – Modeling of reaction times for speech in clear and in noise.

Predictors	Clear					Noise				
	Estimate	*SE*	*CI*	*t*	*p*	*Estimate*	*SE*	*CI*	*t*	*p*
(Intercept)	6.12	.04	6.04–6.21	140.60	<.001***	6.45	.05	6.36–6.54	141.06	<.001***
ZipfF	−.06	.02	−.09–−.02	−3.14	.002**	−.07	.02	−.11–−.04	−4.38	<.001***
FWND	.01	.02	−.03–.04	.39	.695	.02	.02	−.01–.05	1.11	.266
TriF	.01	.02	−.03–.05	.48	.628	−.02	.02	−.05–.02	−1.02	.310
SNR[H.1]						.09	.05	−.01–.20	1.75	.080
SNR[H.2]						.03	.03	−.03–.09	.87	.384
TriF x FWND	−.02	.02	−.06–.01	−1.42	.156	−.03	.02	−.06–−.00	−2.15	.031*
TriF x ZipfF	.01	.02	−.04–.05	.33	.740	.01	.02	−.04–.05	.26	.793
FWND x ZipfF	.03	.02	−.00–.07	1.79	.074	.02	.02	−.01–.05	1.31	.191
TriF x SNR[H.1]						−.02	.01	−.04–.00	−1.77	.077
TriF x SNR[H.2]						−.01	.01	−.02–.00	−1.71	.088
FWND x SNR[H.1]						.01	.01	−.01–.03	.88	.380
FWND x SNR[H.2]						.00	.01	−.01–.01	.43	.666
ZipfF x SNR[H.1]						.02	.01	.00–.04	2.45	.014*
ZipfF x SNR[H.2]						−.01	.01	−.02–.01	−1.02	.306
TriF x FWND x ZipfF	.04	.02	.00–.07	2.06	.039*	.04	.02	.00–.07	2.06	.039*
TriF x FWND x SNR[H.1]						−.01	.01	−.03–.00	−1.67	.094
TriF x FWND x SNR[H.2]						−.01	.01	−.02–.00	−1.71	.088
TriF x ZipfF x SNR[H.1]						.02	.01	−.01–.04	1.22	.221
TriF x ZipfF x SNR[H.2]						.02	.01	.00–.03	2.24	.025*
FWND x ZipfF x SNR[H.1]						−.01	.01	−.02–.01	−.70	.486
FWND x ZipfF x SNR[H.2]						−.00	.01	−.01–.01	−.57	.570
TriF x FWND x ZipfF x SNR[H.1]						−.00	.01	−.02–.02	−.08	.938
TriF x FWND x ZipfF x SNR[H.2]						.00	.01	−.01–.01	.33	.742

Note. Model formular for clear trials: lmer[logRT ~ TriF * FWND * ZipfF + (1 |Participant) + (1 | Word), df.clear, REML = FALSE, control = lmerControl (calc.derivs = TRUE)]; Model formular for noise trials: lmer[logRT ~ TriF * FWND * ZipfF * SNR + (1 | Participant) + (1 | Word), df.noise, REML = FALSE, control = lmerControl(calc.derivs = TRUE)]. SNR [H.1] compared the mean accuracy of SNR+6dB against SNR+10dB, SNR [H.2] compared the mean accuracy of SNR+2dB to SNR+6dB and SNR+10dB *** denotes $p < .001$,** denotes $p < .01$, * denotes $p < .05$.

Table 5 – Modeling the ZipfF-effects on ERP amplitudes for speech in noise.

Predictors	Cluster 1: N400 (440–510 msec)					Cluster 2: P600 (700–770 msec)				
	Estimate	*SE*	*CI*	*t*	*p*	*Estimate*	*SE*	*CI*	*t*	*p*
(Intercept)	−36.76	4.19	−44.97–−28.55	−8.78	<.001***	4.58	4.18	−3.61–12.77	1.10	.273
Baseline	−.06	.02	−.11–−.01	−2.46	.014*	−.14	.02	−.19–−.09	−5.83	<.001***
ZipfF	−3.77	1.39	−6.50–−1.04	−2.70	.007**	5.87	1.99	1.97–9.76	2.95	.003**
SNR[H.1]	2.08	5.08	−7.87–12.03	.41	.682	−6.75	4.80	−16.16–2.66	−1.41	.160
SNR[H.2]	1.06	2.96	−4.73–6.86	.36	.719	−5.56	2.81	−11.06–−.06	−1.98	.048*
Baseline x ZipfF	.02	.02	−.02–.06	.92	.356	.01	.02	−.04–.05	.29	.770
Baseline x SNR[H.1]	.09	.03	.03–.14	3.05	.002**	.03	.03	−.03–.08	.97	.333
Baseline x SNR[H.2]	−.01	.02	−.04–.03	−.44	.662	.02	.02	−.01–.06	1.40	.162
ZipfF x SNR[H.1]	.02	1.56	−3.03–3.08	.02	.987	.82	1.68	−2.47–4.12	.49	.624
ZipfF x SNR[H.2]	.18	.96	−1.71–2.06	.18	.856	.84	1.05	−1.21–2.89	.80	.423
Baseline x ZipfF x SNR[H.1]	.01	.03	−.04–.07	.43	.668	.02	.03	−.04–.07	.61	.545
Baseline x ZipfF x SNR[H.2]	.02	.02	−.01–.05	1.33	.182	.01	.02	−.02–.04	.62	.538

Note. *SNR [H.1] compared the mean accuracy of SNR + 6dB against SNR + 10dB, SNR [H.2] compared the mean accuracy of SNR + 2dB to SNR + 6dB and SNR + 10dB *** denotes $p < .001$, ** denotes $p < .01$, * denotes $p < .05$.*

recognized faster when they resided in dense phonological neighborhoods (Fig. 2A), especially when triphone frequency was also high (Fig. 2B).

When recognizing words in the presence of background noise, triphone frequency had variable effects: While high *TriF* sped up the recognition process of infrequent words at difficult signal-to-noise ratios (+ 2 dB, Fig. 3D), it slowed down RTs at easier SNRs (+ 10 dB, Fig. 3D).

3.3. *EEG: cluster-based permutation*

In order to be able to fit observer-unbiased mixed-models to our EEG data, we chose a data-driven approach to determine regions and times of interest. To this end, we ran six cluster-based permutation tests, one for each predictor (*ZipfF*, *FWND*, *TriF*) in each of the two listening conditions (clear, noise). We found two clusters that suggested significant differences in ERP amplitude

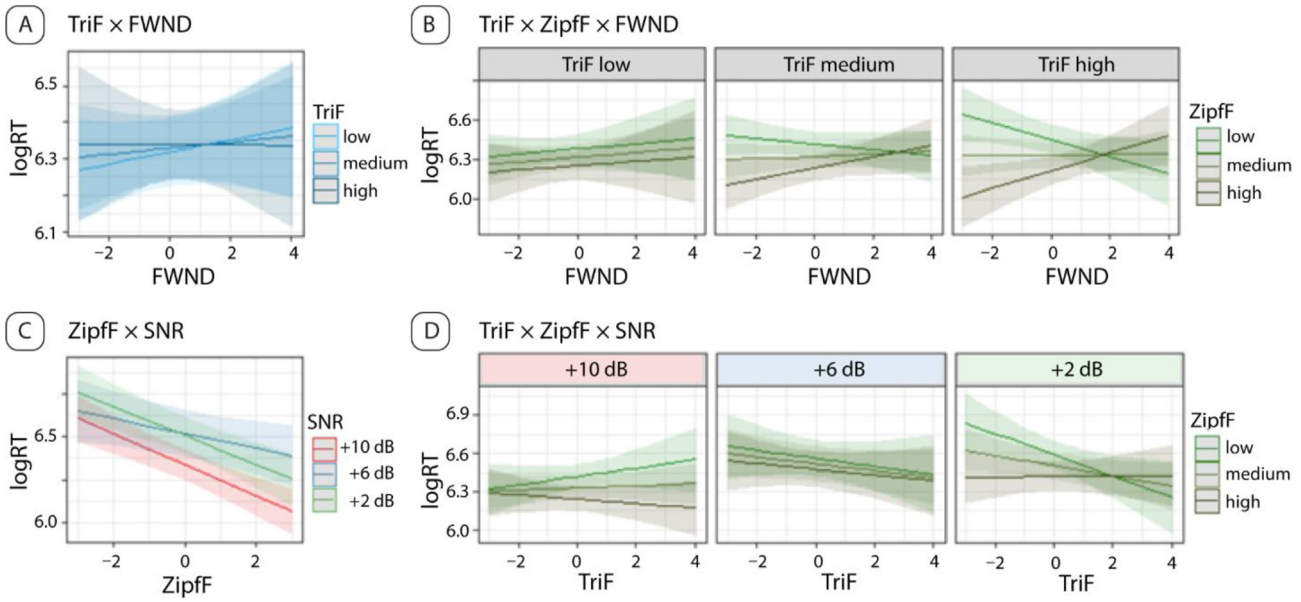

Fig. 3 – Mixed-effects model of reaction times for speech in noise. For illustration purposes only, the continuous variables ZipfF and TriF were split into low (M < −1 SD), medium (M ±1 SD), and high (M > +1 SD). Shaded areas represent the 95% confidence interval. A. Illustration of the interaction of phonotactic probability (TriF) and frequency-weighted neighborhood density (FWND). B. Illustration of the three-way interaction of TriF, ZipfF, and FWND: When TriF was low, FWND did not modulate RTs. When TriF was high, infrequent words profited strongly from high neighborhood density, while the inverse was true for high-frequency words. C. Illustration of the interaction of lexical frequency (ZipfF) and signal-to-noise ratio (SNR): Generally, higher ZipfF led to faster RTs. However, this benefit was weaker at an intermediate SNR of +6 dB. B. Illustration of the three-way interaction of TriF, ZipfF, and SNR. For infrequent words, high TriF slowed RTs at +10 dB, while at +2 dB high TriF was beneficial.

Table 6 – Modeling the FWND-effects on ERP amplitudes for speech in noise and in the clear.

Predictors	Cluster 3: N400 (510–590 msec) in noise					Cluster 4: P200 (260–340 msec) in the clear				
	Estimate	*SE*	*CI*	*t*	*p*	*Estimate*	*SE*	*CI*	*t*	*p*
(Intercept)	−27.91	3.70	−35.16–−20.66	−7.54	<.001***	8.94	5.20	−1.25–19.13	1.72	.086
Baseline	−.09	.02	−.13–−.04	−3.79	<.001***	.09	.02	.05–.13	4.28	<.001***
FWND	−5.05	1.45	−7.88–−2.21	−3.49	<.001***	5.22	1.53	2.22–8.23	3.40	.001**
SNR[H.1]	−1.18	4.50	−10.00–7.63	−.26	.793					
SNR[H.2]	−.44	2.68	−5.70–4.82	−.16	.869					
Baseline x FWND	.01	.02	−.03–.05	.47	.637	−.02	.02	−.06–.02	−1.04	.301
Baseline x SNR[H.1]	.01	.03	−.04–.07	.47	.641					
Baseline x SNR[H.2]	.02	.02	−.01–.06	1.30	.193					
FWND x SNR[H.1]	1.63	1.73	−1.77–5.02	.94	.348					
FWND x SNR[H.2]	−.08	1.16	−2.35–2.18	−.07	.943					
Baseline x FWND x SNR[H.1]	.00	.03	−.05–.05	.07	.947					
Baseline x FWND x SNR[H.2]	−.01	.02	−.05–.02	−.83	.407					

Note. *SNR [H.1] compared the mean accuracy of SNR + 6dB against SNR + 10dB, SNR [H.2] compared the mean accuracy of SNR + 2dB to SNR + 6dB and SNR + 10dB *** denotes p < .001, ** denotes p < .01, * denotes p < .05.*

due to lexical frequency (Fig. 4A1, Clusters 1 and 2) and two that suggested significant differences due to neighborhood density (Fig. 4B and D, Clusters 3 and 5). Furthermore, one cluster was found that suggested a trend for triphone frequency ($p = .054$, Fig. 4C, Cluster 4). We included this trend-level cluster, since it was in line with our hypotheses.

We found that in noise the N400 amplitude was negatively correlated with *ZipfF* over left-parietal electrodes ($T_{sum} = -1778.6$, $p = .047$; see Fig. 4A1) meaning that the higher lexical frequency, the lower (i.e., more negative) the N400 amplitude. Regression coefficients were lowest from 440 to 510 msec post word onset (electrodes: C3, CP5, CP3, CP1, P5, P3, P1). Furthermore, we found that in noise the P600 amplitude was positively correlated with *ZipfF*, again over left-parietal electrodes ($T_{sum} = 2844.6$, $p = .010$; see Fig. 4A2) meaning that the higher lexical frequency, the higher (i.e., more positive) the P600 amplitude. Regression coefficients were highest from 700 to 770 msec post word onset (electrodes: CP3, CP1, P3, P1). In sum,

Table 7 – Modeling the TriF-effect on ERP amplitudes for clear speech.

Predictors	Cluster 5: P200 (240–340)				
	Estimate	*SE*	*CI*	*t*	*p*
(Intercept)	6.58	5.12	−3.45–16.62	1.29	.198
Baseline	.09	.02	.06–.13	4.80	<.001***
TriF	4.04	1.43	1.23–6.84	2.82	.005**
Baseline x TriF	−.04	.02	−.08–−.00	−2.17	.030*

Note. *** *denotes* $p < .001$, ** *denotes* $p < .01$, * *denotes* $p < .05$.

showed the reverse (i.e., positive) directionality with larger P200 amplitudes for words residing in dense rather than sparse neighborhoods (facilitatory neighborhood density effect).

Finally, in clear speech, we found an effect of *TriF* on the P200 ($T_{sum} = 1793.8$, $p = .0549$, 240–340 msec, electrodes: CZ, CPZ, CP2; see Fig. 4C). This effect showed remarkable similarity to the *FWND* effect on the P200 in clear speech and suggested that words with higher *TriF* values elicited larger (i.e., more positive) P200 amplitudes.

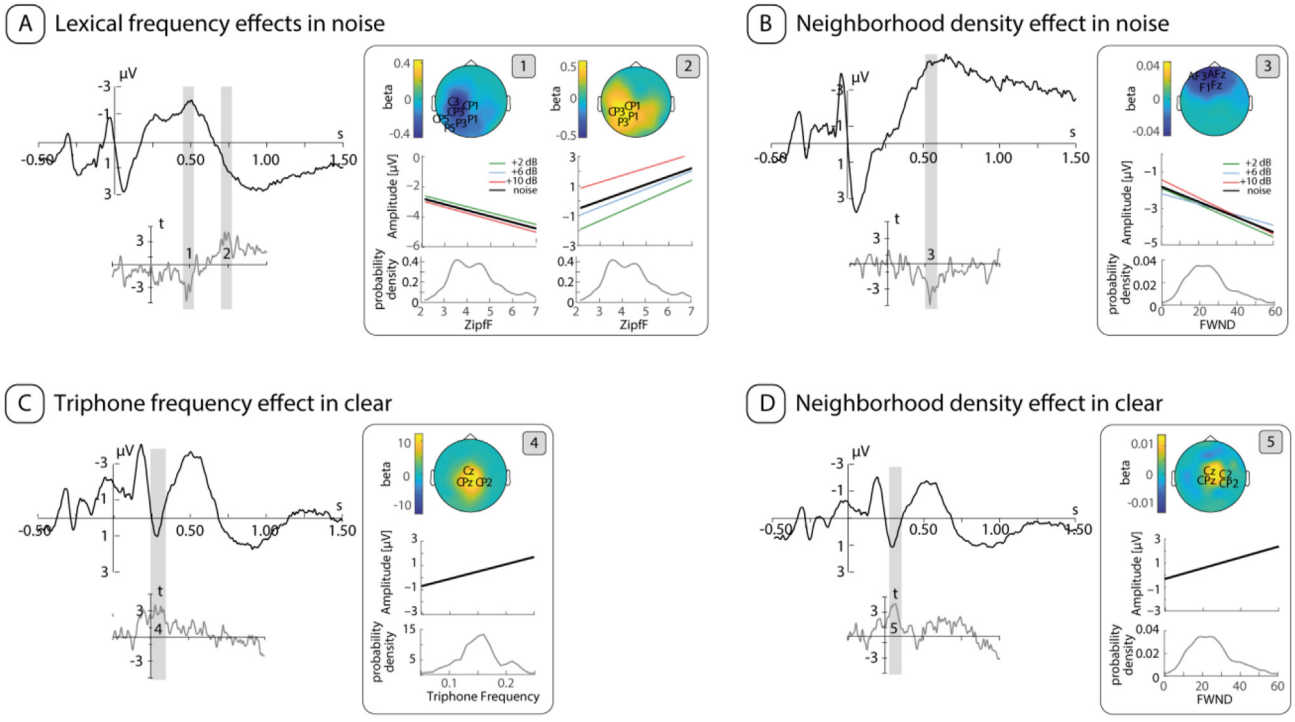

Fig. 4 – Summary of the cluster-based permutation tests. Grand average ERPs are shown for the respective acoustic condition averaged over the region of interest as determined by the respective cluster. The gray shadow marks the significant time window in the ERPs as well as in the line plot of t-values over time. The panel on the right shows the topographies of the regression coefficients (betas) during the determined time of interest. Below the mean least squares linear fits across subjects are plotted aligned with the probability density distribution for the independent variable (*ZipfF*, *FWND*, or *TriF*). A. Two left parietal clusters were found where lexical frequency (*ZipfF*) modulated the processing of words in noise (but not in clear). B. One frontal cluster, within the N400 time window, was found to be modulated by neighborhood density (*FWND*) in noise. C. One central cluster was found, where *TriF* modulated the P200 in clear speech. D. Additionally, the same central cluster was found to be modulated by neighborhood density (*FWND*) in clear speech.

the N400–P600 complex was modulated more strongly in high- compared to low-frequency words.

FWND showed effects in both clear and noisy listening conditions. In noise, *FWND* had an effect on the N400 component from 510 to 590 msec ($T_{sum} = -2208.6$, $p = .019$, electrodes: AF3, AFZ, F1, FZ, Fig. 4B), with larger (i.e., more negative) N400 amplitudes for words residing in dense compared to sparse phonological neighborhoods (inhibitory neighborhood density effect). In clear speech, *FWND* had an early effect on the P200 from 260 to 340 msec ($T_{sum} = 2606.6$, $p = .015$, electrodes: CZ, C2, CPZ, CP2; see Fig. 4D). This effect

3.4. *EEG: linear mixed effects modelling*

Following up on the cluster-based permutation tests, we extracted the EEG data averaged over the determined times and regions of interest. Paralleling the analyses of the behavioral data, we conducted linear mixed-effects models on ERPs. As recommended by Alday (2019), the EEG baseline of each trial (−200 to 0 sec before word onset) was included as a covariate in the mixed-effects model. In all models, significant main effects of baseline were observed. This is to be expected and simply suggests that any ERP amplitude after target onset

is contingent on (correlates with) the amplitude recorded during the baseline period. Note that the baseline in the noise conditions contained the ramping noise before target word onset. We therefore expected the baseline amplitudes to differ in response to different SNRs (reflected as a significant interaction between baseline and SNR).

In line with the cluster-based analysis, we found a negative effect of *ZipfF* on the N400 amplitude ($\beta = -3.77$, SE = 1.39, $t = -2.70$) and a positive effect of *ZipfF* on the P600 amplitude ($\beta = 5.86$, SE = 1.99, $t = 2.95$) in noise. Additionally, we found the P600 amplitude to decrease with noise levels [*SNR(H.1)*: $\beta = -6.75$, SE = 4.80, $t = -1.40$; *SNR(H.2)*: $\beta = -5.56$, SE = 2.81, $t = -1.98$; see also least square fits of the second cluster in Fig. 4A]. No interaction was found between lexical frequency and *SNR* levels (for a summary of results see Table 5).

Furthermore as summarized in Table 6, we confirmed the early *FWND* effect on the P200 amplitude in clear speech ($\beta = 5.22$, SE = 1.53, $t = 3.41$) and the later *FWND* effect on the N400 amplitude in noise conditions ($\beta = -5.04$, SE = 1.45, $t = -3.50$). Again, no interaction was found on the N400 amplitude between *FWND* and *SNR* levels suggesting a uniform negative correlation of *FWND* with the N400 amplitude in noise.

Finally, we confirmed the positive correlation of the P200 with *TriF* in the clear listening condition ($\beta = 4.04$, SE = 1.43, $t = 2.82$; see Table 7). In contrast to the previous models, we observed that baseline EEG amplitude interacted with *TriF* ($\beta = -.04$). This result was unexpected and suggests that the combination of *TriF* and baseline modulated the amplitude of the P200—most likely a spurious finding. Importantly, the presence of this effect did not affect the main effect of *TriF* on the P200 amplitude.

4. Discussion

The present study investigated the interplay between lexical and sublexical processing during spoken-word recognition in the presence of background noise and in clear. We examined the contributions of three predictors (i.e., word frequency, phonological neighborhood density, and phonotactic probability). These predictors have loci at lexical and sublexical levels. We asked whether these predictors explain variation in behavioral indicators of spoken-word recognition performance in both listening conditions. Moreover, we charted the neurobiological markers associated with these properties and examined how listening conditions affected their realization. We tested the hypothesis that listeners adjust their mode of processing and rely more strongly on sublexical processing when the speech signal is less reliable due to the presence of background compared to when it is clear (i.e., absence of background noise). Such a shift should be reflected in the emergence of effects, and/or in an increase in effect size of predictors with a sublexical locus (i.e., phonotactic probability and phonological neighborhood density).

Behavioral findings: The dominance of word frequency and the complementary roles of phonotactic probability and phonological neighborhood density. In line with numerous prior studies, we found word frequency to be the strongest and most stable predictor of word recognition performance (Brysbaert et al., 2018, for review). That is, we observed main effects (better performance for high-versus low-frequency words) of word frequency in clear and in noise for both recognition accuracy and speed. These effects signal a consistent reliance on lexical processing both in clear and in noise.

Phonotactic probability and neighborhood density played a role only in interactions with other predictors. With regard to accuracy, we observed that irrespective of listening condition, phonotactic probability interacted with word frequency such that low-frequency words were recognized more accurately when they were made up of frequent rather than infrequent phonemes (Fig. 1A). Importantly, this benefit was further moderated by listening condition such that as speech quality decreased, the benefit that low-frequency words gained through high-frequency phoneme combinations increased (Fig. 1C). A similar pattern was found for reaction times: In the presence of background noise at an SNR of +2 dB, responses to low-frequency words were faster when composed of frequent rather than infrequent phonemes (Fig. 3D). Importantly, this effect had the opposite directionality in the RT pattern found when words were masked at an SNR of +10 dB, suggesting a change in the contribution of phonotactic probability in response to decreasing speech quality, that is, phonotactic probability became less important when listening conditions were better. Generally speaking, the interaction between word frequency, phonotactic probability and listening condition (SNR) is in line with the notion that sublexical processing is weighted more strongly under conditions of signal degradation (Mattys et al., 2009). In fact, the present results extend the account put forward by Mattys et al. (2009) by demonstrating a stronger weighting of sublexical processing in noise in a task that did not capitalize on acoustic-allophonic variation. Taken together, while lexical information (i.e., frequency) appears to play a mandatory role in spoken-word recognition (i.e., both in clear and in noise), phonotactic probability may provide a 'helping hand' in cases where the speech signal is less reliable (masked).

As discussed in the Introduction, depending on the situational context in which word recognition takes place, neighborhood density has been found to yield both inhibitory (Luce & Pisoni, 1998) and facilitatory (Vitevitch, 2003) effects. This asymmetry has been linked to neighborhood density uniting both lexical (i.e., number of similar sounding words) as well as sublexical (i.e., phoneme probability) elements, which are often positively correlated. The results of the present experiment add to the notion that neighborhood density might indeed be best conceived of as a hybrid measure, with loci at lexical and sublexical levels. In all three models (accuracy, RT_{clear}, RT_{noise}), we observed an interaction between neighborhood density, word frequency and phonotactic probability. For accuracy, the interaction suggested that low-frequency words composed of low-frequency phonemes were recognized less accurately when they had high rather than low numbers of phonological neighbors (i.e., an inhibitory effect). For RTs (both in clear and noise), we observed interesting cross-over patterns, associated with the words' phonotactic probability: While low phonotactic probability resulted in the canonical inhibitory effects (slower RTs for words with more compared to fewer neighbors, uniform across low-, medium-

and high-frequency words), increases in phonotactic probability affected dense- and sparse-neighborhood words differently. Specifically, when phonotactic probability was high, low-frequency words were recognized faster when they resided in dense rather than sparse neighborhoods (i.e., a facilitatory effect). In contrast, when phonotactic probability was high, high-frequency words were recognized faster when they resided in sparse rather than dense neighborhoods (i.e., an inhibitory effect). These effects were not moderated by listening condition. Unlike in previous work (e.g., Luce & Pisoni, 1998), we also did not find main effects of neighborhood density. Taken together, these behavioral results suggest only a minor role of neighborhood density during spoken-word recognition. Importantly, the exact nature of behavioral neighborhood density effects appears to be determined by other lexical and sublexical variables that neighborhood density interacts with. That is, when lexical but not sublexical information provides useful cues (i.e., high word frequency), neighborhood density appears to have inhibitory effects. Conversely, when sublexical but not lexical information provides useful cues (i.e., high phonotactic probability), neighborhood density appears to have facilitatory effects.

In sum, the results render a picture of complex interactions between lexical and sublexical processing subserving successful spoken-word recognition in the clear and in noise. We have visualized these interactions in Fig. 5. Scenarios A, B and C in Fig. 5 embody the accuracy results of the present study. Each panel includes phonemic (i.e., sublexical) and lexical representations and the connections between them. In each panel, we visualized hypothetical variations in sublexical and lexical representations and how they facilitate word recognition. Scenario A represents the main effect of ZipfF and the dominance of lexical frequency during spoken word recognition, which applied to both clear-speech and noisy-speech processing. Similarly, scenario C represents the interaction between lexical frequency, phonological neighborhood density and phonotactic probability, which we observed across clear and noisy listening conditions: Recognition accuracy for low-frequency, low-phonotactic probability targets was highest when they resided in sparse rather than dense phonological neighborhoods. Scenarios B.1 and B.2 highlight the growing importance of sublexical phonotactic probability in noisy listening conditions. That is, while we observed interactions between ZipfF and TriF in clear, the interaction was further moderated by SNR suggesting an increasing influence of sublexical processing for words whose lexical frequency was low.

Electrophysiological findings: Parallel processing of word frequency and phonological neighborhood density in noise. In noise, we observed that words with a higher frequency elicited more pronounced N400 and P600 amplitudes (over centro-parietal electrodes) than words with a lower frequency. These findings are surprising since they suggest that high-frequency words resulted in enhanced processing efforts. We had predicted the opposite directionality (cf. Dufour et al., 2013 but see Strauß et al., 2013). The EEG results are even

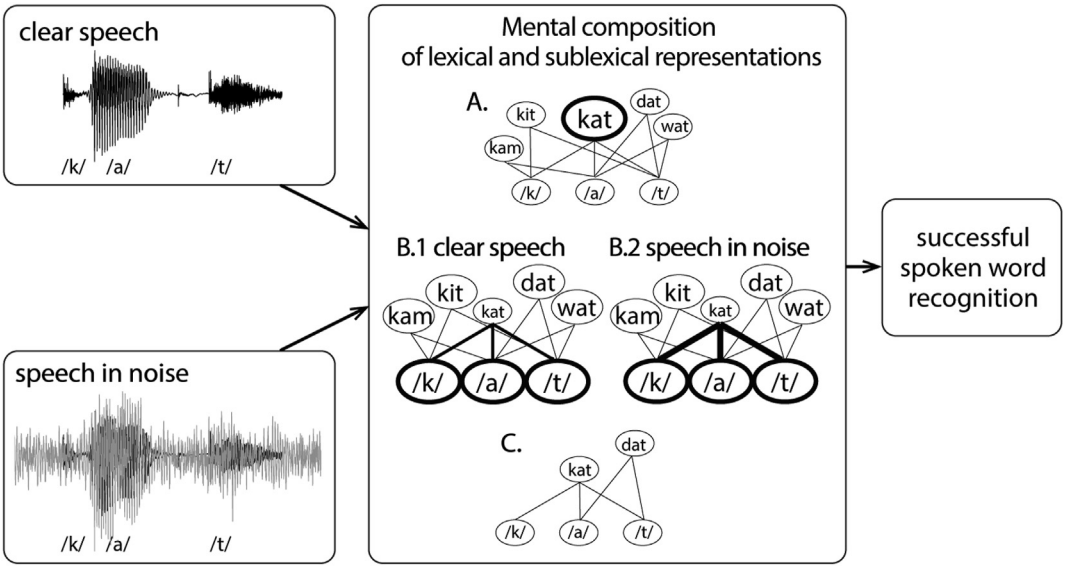

Fig. 5 – Schematic visualization of the mechanisms contributing to successful spoken word recognition in clear and noisy environmental settings. The larger box embodies the interactions between sublexical and lexical processing, with phonemic (i.e., sublexical) representations being connected to lexical representations. Scenario A visualizes the simple effect of lexical frequency in clear and in noisy listening conditions: Thick lines around lexical representations depict high lexical frequency. In scenarios B.1 and B.2, we visualize the increasing importance of phonotactic probability in the presence of background noise. Thick lines around sublexical representations depict high phonotactic probability, which has a positive effect on word recognition when lexical frequency is low. These positive effects are illustrated by thick lines connecting sublexical and lexical representations. Compared to scenario B1, these lines are even thicker in scenario B2, reflecting the interaction with SNR. Finally, scenario C visualizes the three-way interaction between phonotactic probability, neighborhood density and lexical frequency: When phonotactic probability and lexical frequency are low, word recognition is likely to be more successful when neighborhood density is low rather than high (see Discussion for further details).

more surprising since – as discussed above – word frequency had the predicted, canonical behavioral effects. As both the N400 and the P600 components showed a reverse frequency effect, these effects appear unlikely to be mere artefacts. We also observed an effect of phonological neighborhood density, which lay in between the TOIs of the N400–P600 frequency effects (reflecting a 'late N400'). This neighborhood density effect showed over frontal electrodes, with ERP amplitudes being more negative for words residing in dense compared to sparse neighborhoods. This effect was in the same (canonical) direction as the N400 effect reported by Dufour et al. (2013), who interpreted it to reflect the ease of word selection, modulated by the number of similar sounding lexical competitors. We did not observe any EEG effects of phonotactic probability in noise.

One account for the frequency and neighborhood density EEG findings is that they relate to the three-way interaction between ZipfF, FWND and TriF on RTs in noise (visualized in Fig. 3B). As can be seen, high-frequency words (with medium to high phonotactic probability) were most strongly modulated by neighborhood density: When neighborhood density was low, high word frequency resulted in fast RTs; when neighborhood density was high, high word frequency resulted in slow RTs. Interestingly, the EEG effects of frequency and neighborhood density showed up in two distinct neuronal networks working in parallel: frequency over left temporal areas and neighborhood density over frontal regions. On the account outlined above, one would assume interactions between the 'frequency and neighborhood density networks' and one may speculate whether such interactions could have ripple effects, in such a way that processing in one network influences processing in the other network. Clearly, more research is needed to (1) replicate the present pattern of EEG results and (2) to home in on the interactions between the networks, possibly using a different method such as magnetic resonance imaging to be able to visualize fiber tract connections.

Electrophysiological findings: Early effects of sublexical information and absence of lexical frequency effects in clear. As a baseline condition, following presentation in noise, we presented all participants with the same words again in clear. In line with the previous literature, we found that neighborhood density and phonotactic probability modulated the P200 (Dufour et al., 2013; Winsler et al., 2018), both originating from auditory cortices as suggested by the central scalp topography (overlapping electrodes: Cz, CPz, and CP2; see Fig. 4C,D). Both properties were positively correlated with the P200 amplitude, such that dense-neighborhood words and words with frequent phoneme sequences elicited larger P200 amplitudes compared to sparse-neighborhood words and words made up of low-frequency phoneme sequences. In line with the notion that the P200 is sensitive to physical properties of a stimulus (Donchin et al., 1978), we interpret these effects as having a sublexical nature, reflecting facilitated recognition for words with highly probable phonemic sequences. Note that on this account the early neighborhood density effect is a phonotactic probability effect in disguise. Since the RT and accuracy effects in noise provided some evidence for the notion that listeners relied more strongly in noise contexts on sublexical than lexical processing, one may speculate whether the modulations of the P200 amplitude in the clear reflect some form of 'carry-over' effect from the noise block. That is, when carrying out the clear block, participants might still be driven towards a sublexical mode of processing (a remnant from their recent experience on the noise trials), where they capitalize more on the sublexical elements contained in the neighborhood density measure such that high neighborhood density values turned out to have facilitatory rather than inhibitory effects.

Surprisingly and unlike previous studies (e.g., Dufour et al., 2013), we did not find any EEG effects of word frequency in the clear condition. The lack of such effects is especially interesting since we found – as in noise – evidence for a strong role of word frequency in participants' behavior. An explanation for this asymmetry could be that the previous exposure to the stimulus words in the noise block could have mitigated differences in word frequency between targets (i.e., temporarily increased low-frequency words' baseline activation levels, Coltheart et al., 2001). Re-activating the same words in the clear block was thus cognitively less effortful.

5. Conclusion

The present results confirm a simple intuition: Listening in noise is harder than in the clear. This was reflected in both listeners' word recognition accuracy and speed. We also replicated earlier research showing that listeners rely on statistical properties, word frequency, phonological neighborhood density, and phonotactic probability, when recognizing spoken words. Importantly, using the present two-stage lexical decision and transcription task, we have shown that there are complex interactions between these three variables and that the importance and weighting of each variable varies as a function of the listening condition within which comprehension takes place. We tested the hypothesis that listeners rely more strongly on sublexical processing under adverse conditions than in the clear. Such a shift in processing would be reflected in the emergence of effects and/or the increase in effect size of variables having a sublexical locus (i.e., phonotactic probability and neighborhood density). Indeed, our behavioral analyses revealed some support for this claim. The second goal of the present study was to chart the neurobiological markers associated with the three statistical properties and how their realization differs across clear and noise listening conditions. We found three time windows in noise and two in the clear-speech condition where variability in listeners' ERP amplitude was correlated with the three statistical properties. Although the direction of some of the EEG effects was the opposite of what we predicted, we provided interpretations that are in line with the notion that listeners rely more strongly on sublexical than lexical processing in noise compared to clear. However, since these interpretations were conceived post-hoc, an important task for future research is to conduct targeted replications of the present EEG effects. In any case, the present study highlights that a combination of behavioral and neurobiological techniques is indispensable for getting to the bottom of the mechanisms underlying spoken-word recognition in clear and in noise, how these are implemented at the brain level and how they play out in behavior.

Author contributions

FH, JMMQ, and OS conceived the idea for the study and designed the experiment. FH collected the data. TW, AS, and FH analyzed the data and wrote the manuscript. OS and JMMQ provided feedback and commented on the manuscript. All authors approved the manuscript.

Open practices

The study in this article earned Open Data and Open Materials badges for transparent practices. Materials and data for the study are available at https://hdl.handle.net/1839/1fbdb007-96c9-40cb-8031-46af2e019f.

Acknowledgements

This research was supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number 276-89-003) awarded to O.S.

REFERENCES

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology, 56*(12), Article e13451.

Audacity Team. (2014). Audacity (R): *Free audio editor and recorder*. https://www.audacityteam.org/.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340–357.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Benkí, J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *The Journal of the Acoustical Society of America, 113*(3), 1689–1705.

Boersma, P. P. G. (2001). *Praat, a system for doing phonetics by computer (Version 5.1.19) [Computer program]*.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*(1), 1–15.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417–430.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 104–119.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). Drc: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*(3), 256–266.

Cooke, M., García Lecumberri, M. L., Barker, J., & Marxer, R. (2019). Lexical frequency effects in English and Spanish word misperceptions. *The Journal of the Acoustical Society of America, 145*(2), EL136–EL141.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience, 21*(10), 1893–1906.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Event-related brain potentials in man* (pp. 349–411). Academic Press.

Dufour, S., Brunellière, A., & Frauenfelder, U. H. (2013). Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive Science, 37*(3), 489–507.

Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., et al. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods, 50*(3), 1285–1307.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*(5), 501–518.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience, 8*(5), 393–402.

Hintz, F., McQueen, J. M., & Scharenborg, O. (2016, September 2). Effects of frequency and neighborhood density on spoken-word recognition in noise: Evidence from perceptual identification in Dutch. [Talk]. In *22nd annual conference on architectures and mechanisms for language processing (AMLaP 2016)*. Bilbabo (ES).

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America, 29*(2), 296–305.

Hunter, C. R. (2013). Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain and Language, 127*(3), 463–474. https://doi.org/10.1016/j.bandl.2013.09.006

Hunter, C. R. (2016). Is the time course of lexical activation and competition in spoken word recognition affected by adult aging? An event-related potential (ERP) study. *Neuropsychologia, 91*, 451–464.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650.

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology, 1999*(52), 3–6.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*(2), 119–131.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.
Lewendon, J., Mortimore, L., & Egan, C. (2020). The phonological mapping (mismatch) negativity: History, inconsistency, and future direction. *Frontiers in Psychology, 11*, 1967.
Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*(3), 615–625.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36.
Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology, 9*, 369.
Magnuson, J. S., You, H., Luthra, S., Li, M., Nam, H., Escabi, M., et al. (2020). Earshot: A minimal neural network model of incremental human speech recognition. *Cognitive Science, 44*(4), Article e12823.
Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *Plos One, 7*(8). https://doi.org/10.1371/Fjournal.pone.0043230
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203–243.
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978.
Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477–500.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.
McQueen, J. M. (2003). The ghost of christmas future: Didn't scrooge learn to be good?: Commentary on Magnuson, McMurray, tanenhaus, and aslin (2003). *Cognitive Science, 27*(5), 795–799.
Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*(2), 165–178.
Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 873–889.
Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language, 64*(4), 460–476.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234.
Norris, D., & McQueen, J. M. (2008). Shortlist B: A bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.
Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(3), 299–325.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*.
Pollack, I., Rubenstein, H., & Decker, L. (1960). Analysis of incorrect responses to an unknown message set. *The Journal of the Acoustical Society of America, 32*(4), 454–457.
Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *The Journal of General Psychology, 13*(1), 199–203.
Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81*(1–3), 666–678.
R Core Team. (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/.
Scharenborg, O., Coumans, J. M., & van Hout, R. (2018). The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(2), 233–249.
Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience, 25*(8), 1383–1395.
Strauß, A., Kotz, S. A., Scharinger, M., & Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage, 97*, 387–395.
Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition, 22*(3), 259–282.
Van Engen, K. J., Dey, A., Runge, N., Spehar, B., Sommers, M. S., & Peelle, J. E. (2020). Effects of age, word frequency, and noise on the time course of spoken word recognition. *Collabra: Psychology, 6*(1). https://doi.org/10.1525/collabra.17247
van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190.
Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition, 18*(4), 380–393.
Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clinical Linguistics & Phonetics, 17*(6), 487–499.
Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325–329.
Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374–408.
Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*(3), 481–487.
Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 75–94.
Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1–2), 306–311.
Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3*(1), 64–73.
Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(3), 387–401.
Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2018). An electrophysiological megastudy of spoken word recognition. *Language, Cognition and Neuroscience, 33*(8), 1063–1082.
Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language, 48*(4), 779–793.