From Sound Waves to Words:
The Effects of Phonotactic Probability on Speech Processing

Alice Shen

University of California – Berkeley

Summer Research Opportunity Program (SROP)
Abstract

Phonotactic probability is the likelihood of sounds and sound sequences occurring in a language. Recent speech perception models have suggested that our brains use more than just acoustic cues during speech processing, but also incorporate phonotactic information. Studies have found that responses to high probability non-words had faster reaction times than low probability non-words in an auditory repetition task. In the present experiment, three participants completed a similar task with 37 word and non-word synthesized stimuli. In contrast to previous research, a linear mixed effects model was used for data analysis instead of dichotomization and ANOVA. Phonotactic probability’s heavily-studied role in learning new vocabulary suggests that findings will have an impact on first and second language acquisition, as well as on aphasia word retrieval therapy.
This study attempts to answer part of the fundamental question: how do human beings understand speech? It specifically focuses on the usage of phonotactic probability in speech processing. The role of phonotactic probability has implications for language acquisition and word learning. For example, the instruction of ESL students will benefit from knowing how phonological representation affects learning, and adults with aphasia can possibly regain vocabulary more efficiently (Freedman & Barlow 2013).

Phonotactic probability is the likelihood of sounds and sound sequences occurring in a language. If the sound /a/ has higher phonotactic probability than /i/ in English, it means that /a/ occurs more often than /i/ in English words. Phonotactic probability is positively correlated with neighborhood density, which measures the number of phonologically similar words to a word (Storkel, Maekawa, & Hoover, 2010). For example, neighbors of the word “can” include “con,” “ban,” and “cat,” all of which differ from “can” by one sound. Therefore, if a word has high phonotactic probability, it also has high neighborhood density.

Interactive models of speech processing such as TRACE and NAM propose that phonological (sub-lexical) and lexical form representations are used in perception (McClelland & Elman, 1986; Luce & Pisoni, 1998; Vitevitch & Luce, 1998, Storkel, Armbruster, & Hogan, 2006). Phonological representations refer to knowledge of specific sounds and parts of the word, while lexical representations refer to knowledge of word as a whole. When we perceive a word that is present in our mental storage of vocabulary (lexicon), corresponding phonological and lexical representations are triggered in memory and used toward recognition of that word. In
contrast, upon perception of an unfamiliar word, phonological representations are activated and matched, but no known lexical representations can be matched (Storkel et al., 2006).

Previous studies examined the effects of phonotactic probability and neighborhood density on processing, and have proposed that they play roles in phonological and lexical levels of processing respectively. To investigate these roles, word and non-word stimuli were employed. Non-word refers to a group of letters or sounds that does not have a meaning and is not accepted by the speakers of a language. “Mabberwocky,” “leference,” and “nord” are examples of non-words in English.

Vitevitch and Luce (1998), in an auditory repetition task, found that responses to high probability-density non-words had faster reaction times than low probability-density non-words, while the opposite effect occurred for words. They concluded that perception of non-words is aided by high phonotactic probability. Since non-words are not stored in the lexicon, only phonological representations are activated and used. They suggested that, when a non-word has high probability, there are more existing representations that can aid that perception and production process, thus resulting in a quicker reaction time. A low probability non-word would therefore not have as many existing representations to aid its processing. On the other hand, perception of words is inhibited by higher neighborhood density, and the presence of many other similar-sounding words causes lexical competition. Lexical competition involves the representations of the perceived item and of similar-sounding words being activated in the individual’s lexicon during perception. The activated neighbors “compete” until finally those hypotheses are whittled down and one is considered the most likely (Luce & Pisoni, 1998). That “winning” hypothesis is what the listener ultimately “hears”, and in an auditory repetition task, is also the subject’s “repetition,” or production. Because high probability words have high
FROM SOUND WAVES TO WORDS

neighborhood density and activate more hypotheses than low probability words, a slower reaction time results. In summary, in high probability-density items, a facilitatory effect (which is attributed to high probability) occurs for non-words, while the lexical status of high-density words dominates, bringing about an inhibitory effect.

The same-different task consists of a participant deciding whether a pair of spoken stimuli is the same or not. Vitevitch and Luce (1999) conducted two experiments using this task, blocking word and non-word stimuli for the first experiment and intermixing stimuli for the second. They obtained the same result as for their auditory repetition task (1998): reaction times to words were inhibited by high probability-density, but reaction times to non-words were facilitated by high probability-density. Non-words still had an overall longer reaction time.

However, tasks other than auditory repetition and same-different have been used to assess the effects of phonotactic probability. One is the lexical decision task, which, unlike auditory repetition tasks, requires a judgment response, which is usually a button press following the visual presentation of the stimulus. Subjects judge the lexicality of the stimulus, meaning they decide whether they think the stimulus is a word or not. Experiments using this task, whether with audio or visual stimuli, have found that subjects responded more slowly to high probability-density stimuli than low probability-density stimuli for words and non-words (Vitevitch & Luce, 1999; Pylkkanen, Stringfellow, & Marantz, 2002). This seems like a contradiction to the result of the auditory repetition task, which made a distinction between the effect’s consequences for words versus non-words. Pylkkanen et al. (2002), explain this by the connotation of the stimulus’s lexicality. In a lexical decision task, the presentation of non-word stimuli necessitates the activation of phonologically similar words from the lexicon in order to judge stimulus
lexicality (Pylkkanen, et al., 2002). Consequently, reaction time would be slower for non-words with high probability-density.

Hence, previous experiments show differentiation between the effects of phonotactic probability and neighborhood density that depends on task context. In auditory repetition and same-different tasks, words and non-words showcase opposite effects because of dominance from density and probability, respectively. Nevertheless, during lexical judgment tasks, lexical processing is core to discrimination between stimuli, which is core to the task, and so the facilitatory effect in high probability-density non-words is reversed by the inhibitory effect of neighborhood density.

The research of Lipinski and Gupta (2004) replicated Vitevitch and Luce (1998), using digitally equalized stimulus durations, under the hypothesis that differences in stimulus duration might have caused the facilitatory effects for high probability-density items. They found that this effect disappeared once stimulus durations were equalized, prompting a response from Vitevitch and Luce (2005). The latter argued that the inclusion of inaccurate responses in analysis, the quick pace of stimulus presentation, and pressure to respond may have obscured the effects of phonotactic probability.

The present study examines the effect of phonotactic probability on reaction time, using an auditory repetition task. In contrast to previous studies, phonotactic probability is not dichotomized and analyzed as a categorical variable with high and low groups. Instead, a linear mixed effects model is used, treating phonotactic probability as a continuous variable, and including lexicality as the other fixed effect. Random effects were also included, for participant and stimulus. These methods were necessary because of the nature of the data collected, which
FROM SOUND WAVES TO WORDS

will be further described. Similar to previous studies, phonotactic probability was evaluated through the measures of position-specific biphone frequency and positional segment frequency. Moreover, word and non-word stimuli were intermixed. Only accurate responses were included in analysis, and participants were instructed to wait two seconds for a visual cue before responding.

I hypothesized that, for reaction time, non-words and word would exhibit respectively the facilitatory probability effect and the inhibitory density effect, while non-words would generally have slower reaction times versus words. In other words, I expect to obtain the same result as the aforementioned studies.

Methods

Participants

The participants in this experiment were three patients from the epilepsy monitoring ward in UCSF Medical Center. They were native English-speaking adults and had no sign of any speech disorder. The data was collected from this sample as part of an experiment on the cortical organization of speech, and neural activation during mismatched perception and production. For these participants, seizures were localized to areas of the brain that do not affect speech, so their medical condition should not affect their speech processing in general (Bouchard, Mesgarani, Johnson, & Chang, 2013).

Materials

A list of 37 monosyllabic CVC (consonant-vowel-consonant) stimuli consisting of words and non-words was used in this experiment. Of those, 15 were non-words, and 22 were words. (Quasi-words such as prefixes were treated as words).
Phonotactic Probability

To determine phonotactic probability, I used an online Phonotactic Probability Calculator (Vitevitch & Luce, 2004). The calculator displayed phonotactic probability for entries based on measures of positional segment frequency and biphone frequency from the Merriam-Webster Pocket Dictionary. Positional segment frequency is how often a single sound segment occurs in a certain position in a word, in a specific language. For example, in English, if the sound /k/ occurs more frequently at the beginning of a word than at the end, then its positional segment frequency is higher for word-initial position. Biphone frequency measures the probability of two sounds occurring next to each other in a specific position in a word in a language. For example, the pairing /k/ and /a/ have a higher biphone frequency than /t/ and /a/ when at the start of an English word.

Instead of dividing stimuli into high and low probability sets for words and non-words, I used phonotactic probability as a continuous variable. The median split method used by Vitevitch & Luce (1998, 1999), categorizes this numerical variable. Phonotactic probability had to be measured separately through biphone frequency and positional segment frequency. Furthermore, as previously stated, the stimulus set was created for the purposes of a neurolinguistics experiment, and thus was not balanced for the purposes of my experiment.

Procedure

The experiment was conducted inside an epilepsy monitoring ward. In the task, the participant heard the stimulus (which was produced by a speech synthesizer) and then repeated back what they heard into a microphone once they received a visual cue (there was a 2-second delay between stimulus onset and onset of the visual cue to start responding). They did this for
37 stimuli in each run of the task, and each participant performed this task eight times over a period of 2-5 days. In each run, the order of the stimuli was varied, so words and non-words were always randomly intermixed. The experimenter informed participants that some stimuli would sound familiar and others unfamiliar, and asked them to repeat back each stimulus exactly as they thought they heard it.

Reaction time was defined from the onset of the stimulus to the onset of the participant’s response. The entire task, from the beginning audio cue, including all stimuli and responses, was recorded. Only accurate responses were used in data analysis. Responses that matched stimuli in all phonetic segments were considered accurate. This was determined through manual transcription of the recording.

**Results**

I ran a linear mixed effects model to study the relationship between phonotactic probability, lexicality, and reaction time. Simultaneously, I can account for non-independence by adding the random effects for participant and stimulus to my model. This is necessary because the task involved 37 stimuli, and each participant completed the task eight times. Thus, each participant contributed anywhere from 200 to 300 responses (inaccurate responses were not counted). Reaction times collected from the same participant cannot be considered independent from each other. Similarly, reaction times from responses to the same stimulus are not independent from each other.

Reaction time was modeled with a linear mixed effects model, with the fixed effects of positional probability and the interaction between biphone probability and lexicality. Random effects for participant and stimulus were also included. With only fixed effects, this model
accounted for 0.6% of variance in the data, whereas including random effects as well, it accounted for 27.6% of variance in the data.

An increase in biphone probability is associated with an increased reaction time ($p < 0.05$). Lexicality and positional segment probability showed no significant effect on reaction time. However, the effect of the interaction between biphone probability and lexicality was near-significant ($Estimate = -16.5318, p = 0.07$): when looking at a word stimulus instead of a nonword stimulus, the magnitude of biphone probability’s effect on reaction time decreases (see Table 1 for full results).

**Table 1**

| Fixed Effect                                      | Estimate | MCMC mean | Lower 95% HPD | Upper 95% HPD | pMCMC   | Pr(>|t|) |
|--------------------------------------------------|----------|-----------|---------------|---------------|---------|---------|
| (Intercept)                                      | 2.6419   | 2.6461    | 1.9258        | 3.3563        | 0.0001  | 0.0000  |
| Biphone Probability                              | 18.7477  | 18.6448   | 0.0550        | 36.5098       | 0.0486  | 0.0227  |
| Lexicality (word)                                | 0.0939   | 0.0897    | -0.0614       | 0.2480        | 0.2470  | 0.1804  |
| Positional Segment Probability                   | -0.1989  | -0.1927   | -1.3909       | 0.8849        | 0.7416  | 0.7000  |
| Interaction of Biphone Probability and Lexicality (word) | -16.5318 | -16.3398 | -36.5826      | 3.4409        | 0.1082  | 0.0726  |

**Discussion**

First, it is important to note that my method differed from those of previous researchers; I did not treat phonotactic probability as a categorical variable by using a median split. Such
dichotomization methods may cause problems, including loss of statistical power and risk of a Type I error, due to the arbitrary cut-off where data items that are actually close are characterized as being very different (Altman & Royston, 2006; MacCallum, Zhang, Preacher, & Rucker, 2002). I chose to keep the phonotactic probability measures as the continuous variables that they inherently are, using linear mixed models. Not only do linear mixed effects models not alter the differences among individual data items, but the inclusion of random as well as fixed effects allows a better fit for the data. Thus, through use of a random effect, the model is able to consider, for example, that there were only three participants who came from a convenience sample (Baayen, Davidson, & Bates, 2008). In addition, a random effect controlled for the naturally unequal stimulus durations that result from stimuli being comprised of different phonemes.

Furthermore, the experimental situation was not ideal, because of the participants’ medical condition and the hospital setting. There were several interruptions during recordings, such as the hospital intercom, that resulted in the participant not hearing the stimulus and either not responding or giving an incorrect response that clearly pertained to the previous stimulus. Accordingly, only accurate responses were included in analysis. Inaccurate items should be analyzed separately, also because they might reasonably be associated with prolonged reaction times, depending on the functional origin of the error. Yet although inaccurate responses were removed prior to analysis, it is clear how those disruptions can affect upcoming reaction times. Again, the data used for this study was originally meant for an experiment that required the participation of neurosurgical patients, which is why each participant did eight repetitions of the task (Leonard, Bouchard, & Chang, 2013). That there are significant findings is remarkable.
FROM SOUND WAVES TO WORDS

The linear mixed model for reaction time reveals that, in general, stimuli with higher biphone probability are associated with slower reaction times. In other words, higher probability words and non-words take longer to process than lower probability words and non-words. There are several possible interpretations of this finding.

One interpretation is that there is no phonotactic probability effect that facilitates the processing of high probability non-words. In fact, “effects of probabilistic phonotactics (even for nonsense words) may arise from interactions among lexical items themselves, implicating no direct role of sublexical units (Vitevitch & Luce, 1998). This could mean that sublexical units are only activated in the context of lexical units. Phonotactic probability could only play a role in processing, in that an item with high probability must therefore have high density, because the probability-density correlation is sufficiently high. Lexical competition among activated phonological neighbors could occur dominantly for both words and non-words, with sublexical units that activate within the lexical unit environment. This would likewise explain why non-words were shown to take longer to process than words; finding a matching lexical unit would reasonably be faster than confirming the non-existence of a lexical unit that contains all necessarily activated sublexical units.

Another possibility is that the phonotactic probability effect is indeed a fragile one that depends on presentation timing (Vitevitch & Luce, 2005). Perhaps the two second window between the stimulus onset and the onset of the visual cue to start responding masked this subtle yet present effect.

Since there is much research conducted with a different method that leans toward the existence of a facilitatory effect of phonotactic probability on non-words, I refrain from making any definitive conclusion, and instead recommend that further research be conducted. Avoiding
dichotomization and using a linear mixed effects model seems advisable, considering the latter’s advantages and that phonotactic probability is inherently a continuous variable (Altman & Royston, 2006; MacCallum et al., 2002). Moreover, a more controlled experimental situation, the use of a random, larger sample, and lack of repeated stimulus items from participants would be ideal for further studying this effect.
## Appendix

### List of Stimuli (in Arpabet)

<table>
<thead>
<tr>
<th>Words</th>
<th>Non-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-AE1-D</td>
<td>D-AE1-Z</td>
</tr>
<tr>
<td>D-IH1-D</td>
<td>D-EH1-Z</td>
</tr>
<tr>
<td>G-AH1-M</td>
<td>G-AH1-TH</td>
</tr>
<tr>
<td>K-AA1-N</td>
<td>K-AA1-S</td>
</tr>
<tr>
<td>K-IY1-N</td>
<td>K-IY1-K</td>
</tr>
<tr>
<td>K-UW1-K</td>
<td>K-IY1-S</td>
</tr>
<tr>
<td>K-UW1-N</td>
<td>K-UW1-S</td>
</tr>
<tr>
<td>M-AH1-M</td>
<td>M-AH1-TH</td>
</tr>
<tr>
<td>N-AA1-K</td>
<td>N-AA1-S</td>
</tr>
<tr>
<td>N-AA1-N</td>
<td>N-IY1-K</td>
</tr>
<tr>
<td>N-IY1-S</td>
<td>N-IY1-N</td>
</tr>
<tr>
<td>N-UW1-K</td>
<td>S-AA1-N</td>
</tr>
<tr>
<td>N-UW1-N</td>
<td>S-UW1-K</td>
</tr>
<tr>
<td>N-UW1-S</td>
<td>S-UW1-S</td>
</tr>
<tr>
<td>S-AA1-K</td>
<td>Z-IH1-D</td>
</tr>
<tr>
<td>S-AA1-S</td>
<td></td>
</tr>
<tr>
<td>S-IY1-K</td>
<td></td>
</tr>
<tr>
<td>S-IY1-N</td>
<td></td>
</tr>
<tr>
<td>S-IY1-S</td>
<td></td>
</tr>
<tr>
<td>S-UW1-N</td>
<td></td>
</tr>
<tr>
<td>TH-AH1-M</td>
<td></td>
</tr>
<tr>
<td>Z-EH1-D</td>
<td></td>
</tr>
</tbody>
</table>
FROM SOUND WAVES TO WORDS

References


FROM SOUND WAVES TO WORDS


FROM SOUND WAVES TO WORDS
