The Effect of Word Familiarity and Lexical Neighbourhood on L2 Speech Perception

Erina Sawada
Sophia University

1 Introduction

There has been a lot of study on the lexical effect on word recognition, contributing to figure out the mechanism of top-down speech processing. Many studies investigated perceiver's performance as a function of some properties of words such as frequency, familiarity, and neighbourhood. Among those lexical properties, frequency is the most widely used variable in speech perception studies (Nusbaum, Pisoni, and Davis, 1984). Scarborough, Cortese and Scarborough (1977), for example, showed that high-frequency words were recognized faster than low-frequency words. They assumed the frequency value stored in the mental lexicon derives from counting each occurence in the language. Since the frequency effect is widely studied and confirmed, some lexical access model also take frequency values into an account. For example, the logogen model presupposed the frequency effect on word recognition. The threshold of each logogen's activation level varies according to its frequency counting (Morton, 1969 1982). High-frequency words have lower threshold than that of low-frequency words, which requires less information to recognize the former than the latter.

1.1 *Familiarity Effect* Although the word frequency effect was evident and was incorporated into many lexical access models, it is questionable whether the frequency of printed words is an adequate variable to describe one's mental lexicon. Gernsbacher (1984), for example, challenged the idea of counting printed word as an estimation of perceivers' lexicon. She suggested that the familiarity rating was a more comprehensible variable than printed words frequency, given that she found familiarity measure was a better predictor of one's reaction time than other lexical properties in her experiments. Despite the fact that there was the correlation between familiarity and frequency in most of the set of 130 words (r^2=0.81), it was not the case for the lowest-frequency words. Thus, she attributed this asymmetric relation to the nature of written word frequency and subjective familiarity. Compared to the frequency, which is incremented based on occurrences in text corpora, the subjective familiarity for a given word is based on all occurrences of one's own production and audio/visual perception. Based on her argument, Connine, Mullennix, Shernoff, and Yellen (1990) conducted four experiments to compare subjective familiarity and printed words frequency using both visual and audio stimuli in a lexical decision task and a naming task. They found familiarity effect across modalities and tasks, which was not the case for frequency effect that varies with modalities and tasks.

Nusbaum et al. (1984) also pointed out the discrepancy between familiarity and frequency. They recruited 600 native American English speakers to collect their familiarity ratings of about 20,000 words. As a result, frequency and familiarity had a weak connection, and the correlation between them only accounted for 17.5% of the variance of familiarity ratings. There was no frequency cutoff for familiarity ratings. In other words, it is possible to rate a word as familiar irrespective of its low frequency value. They attributed this discrepancy to the sampling error of frequency since the frequency counting could not take word occurrence in casual speech into an account. Due to the limitation to calculate frequency values, word frequency is not be equal to speakers' subjective familiarity, as suggested in Gernsbacher (1984). Interestingly, Nusbaum et al. (1984) also reported that subjective familiarity value did not correlate with word length (r^2=0.04, ns), unlike the strong correlation between frequency and word length.

1.2 *Neighbourhood Effect* Lexical effects are not only limited to the properties of a particular word itself. Rather, similar words to a given word, called neighbours, also have an influence on speech perception. For example, in late 1960s and early 70s, the frequency effect on speech perception has been studied, and those studies assumed that phonological make-up of words is perceptually equivalent between high and low frequency

* I would like to thank my advisor Professor Mafuyu Kitahara for encouraging me to apply for the Asian Junior Linguistic Conference, and all my participants for taking part in my experiment. I am also grateful to AJL committee for their efforts to organize the conference and to edit the proceedings.

conditions.

Landauer and Streeter (1973) pointed out the structural difference between high-frequency words and low-frequency words, however. Following their argument, Luce (1985) analyzed the distributional difference of common (more frequent) and rare (less frequent) words in Webster's Pocket Dictionary in relation to frequency of the words themselves, mean frequency of neighbours, number of the neighbours, and phonemic configuration (see however Pisoni, Nusbaum, Luce, and Slowiaczek, 1985). He found that the mean frequency of neighbours of high frequent words were higher than that of low frequent words when the range of word length was two to four phonemes, even though the number of neighbours in both conditions were approximately equal. He indicated that the highly frequent words were more likely to be confused than less frequent words in that high frequency words tend to have high frequent neighbours more than low frequent neighbours. In addition to his analysis of neighbourhoods and its density, he also suggested that the frequency effects given in the previous studies were due to the difference of the structure between high and low frequent words, rather than frequency value itself (see however Pisoni et al., 1985).

Although definitions of neighbourhood may vary, the most basic and widely used one in the psycholinguistic literature is such that a group of words differing from a particular word by just one segment (see Pisoni et al., 1985; Greenberg and Jenkins, 1964). A word *pat*, for example, may have neighbours such as *past, at, put, pit, pad* by adding, deleting, or substituting a single segment. In Greenberg and Jenkins' study, a set of nonce words varying in a number of neighbouring real words in English was created. They conducted various experiments to investigate possible measurement options of the distance between nonce words and real words in English. They found that the number of English words created by substituting a single segment from the original one was highly correlated with two subjective evaluations of the distance between nonce words and English words: free magnitude estimation and an 11-point rating scale (Greenberg and Jenkins, 1964: p165).

Luce and Pisoni (1998) proposed a model of lexicon based on the Greenberg and Jenkins' neighbourhood concept combined with the choice rule by Luce (1959), which is named as Neighbourhood Activation Model (NAM). They used frequency values of words and their neighbours to categorize four types of distributions. The following schematic diagram (Figure 1) summarizes their characteristics. Four bars in the center represent particular target words, and surrounding thin bars represent neighbours of the target. The height of the bar corresponds with the token frequency of each word.

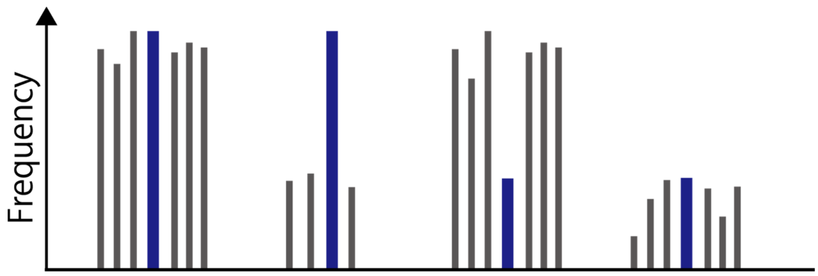

Figure 1: Schematic Diagram of Categorization from Neighbourhood Activation Model

More recently, Vitevitch and Rodriguez (2005) reported the neighbourhood effect appeared in the high-frequency word condition, but not in the low-frequency word condition, in an L1 lexicality judgment task in Spanish. They also found that the frequency of a particular word and that of neighbors were not independent of each other, rather they had a combined effect on speech perception. In the high-frequency condition, the reaction time to the words with high-frequency neighbours was shorter than those in the low-frequency neighbourhood condition. They noted that the length of words correlated negatively with the number of neighbours, that is, shorter words had more neighbours than longer words. Their results agreed with the study by Pisoni et al. (1985) in English.

Although many studies on lexical effects have been conducted for about five decades, little of the mechanism and development of word recognition in L2 has been probed in experimental studies. For example, the study by Van Heuven, Dijkstra and Grainger (1998) gave an important insight to the construction of lexicon not only in L1 but also in L2. Van Heuven et al. (1998) conducted a crosslinguistic lexicality judgment task; the bilinguals of Dutch and English judged English words, which had orthographic neighbours in Dutch. They found a crosslinguistic inhibitory effect of neighbourhood on the task where the participant responded to the words with more neighbours in longer reaction time. They concluded that bilinguals might have a single mental lexicon with some items shared at least visually in two languages rather than two discrete lexica. Although their study proposed the form of the bilinguals' lexicon, it remains uncertain about the mechanism of speech perception and its interaction with lexical effects among L2 learners. Considering the scarcity of research in L2 learning, the present study asks three research questions as follows: (1) Do learners get influenced by neighborhood and familiarity measured in the target language? (2) If so, how the property of neighbourhood and familiarity in the target language makes difference in their performance? And lastly, (3) Within L2 learners, does the proficiency level have any effect on their performances with respect to the lexical property of words? Reasons for using familiarity measure rather than frequency-based measure will be discussed in section 2.2 Target and Fillers.

2 Method

2.1 *Participants* The present study collected data from 24 native speakers of Japanese who learn English as their second language (*age:* 19-53, *Male*=14, *Female*=10). All of the participants were from two universities in Tokyo. They were asked to fill the questionnaire on their language backgrounds and other demographic information. In the questionnaire, they also reported their scores of various standardized English tests. Based on their scores and the table mapping those standardized scores to CEFR levels in MEXT (2018), participants were categorized into two groups: intermediate (*N*=12) and advanced learners (*N*=12). Advanced learners reside from B2 up to C1 level in CEFR; intermediate learners are from A2 to B1 level. Table 1 summarizes the breakdown of the number of participants in each level.

CEFR	*N*	*Categories*
C2	0	-
C1	8	[Advanced = 12]
B2	4	
B1	9	[Intermediate = 12]
A2	3	
A1	0	-

Table 1: English Proficiency of the Participants (*N*=24)

2.2 *Targets and Fillers* The present study used familiarity rating of a word and its neighbours as independent variables. As mentioned in the introduction, word frequency has been widely used in L1 study; however, this measure is not adequate for investigating the mental lexicon. Although word frequency and familiarity are correlated more or less, many scholars (Gernsbacher 1984; Nusbaum et al. 1984; Connine et al., 1990) suggested that familiarity was better than frequency in terms of the adequacy of their account for the performance in tasks. The discrepancy between familiarity and frequency could be more crucial when a study investigates learners' speech perception. Since the learners have more limited access to their target language than native speakers, word frequencies, whose calculation is based on texts for native speakers, cannot be directly applicable to the net input for L2 learners. Such learners, of course, may have a different set of word familiarity values as well, still familiarity can be used as a better analogue than word frequency for their internalized lexical knowledge based on both production and perception, as suggested in Gernsbacher (1984)[1].

Following NAM by Luce and Pisoni, the schematic diagram below illustrates four types of lexical items. As opposed to their model using word frequency, the present study used word familiarity for the reasons mentioned above. Two of the four categories have highly familiar words (hereafter *targets*), and one of which has neighbours with high familiarity (named as FOREST) and the other has neighbours with low familiarity (POLE). Likewise, another two categories have low familiar targets, but they are different in their neighbours' familiarity values: high (CONCAVE) and low (BUSH). The present study uses those names for each word category for the sake of convenience. Their names are based on visual analogies of the diagram below.

[1] To confirm that the stimuli used in the present experiment have no confounding relation between those frequency and familiarity, each word's lexical properties including familiarity, frequency, neighbourhood familiarity, neighbourhood frequency, and neighbourhood density were extracted from Hoosier Mental Lexicon and utilized to conduct a post hoc model test. The test revealed that a model including familiarity value of each stimulus (RT as dependent variable and familiarity of the target, mean familiarity of the neighbours, and their density as independent variable, AIC=4523.3) had a better fit than a model including log transformed frequency value (AIC=4624.9) at 0.001 significance level.

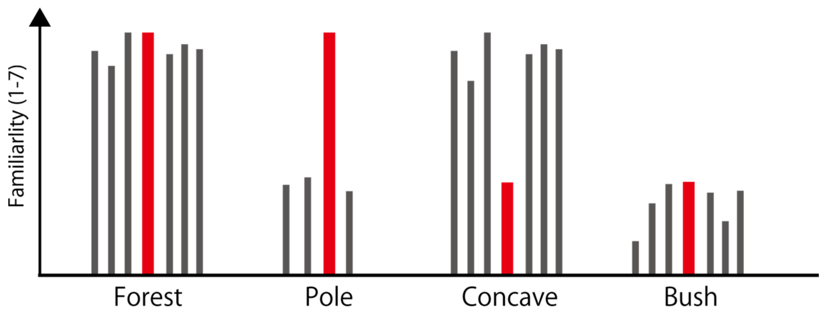

Figure 2: Schematic Diagram of Four Lexical Types Based on Neighbourhood and Familiarity

The thick bar at the center of each category shows a particular target. The height of each bar corresponds to its familiarity. The thin bars surrounding target bar represents neighbours. Table 2 shows the targets and fillers used in the present study, sorted according to their word familiarity and mean familiarity of neighbours. For the sake of convenience, hereafter, this paper refer to the four types of combination of word familiarity and mean familiarity of neighbours as *neighbourhood types*.

Categories	*Targets*	*Fillers*
FOREST	joke, age	coast, fault, foot, hut, sting, pound
POLE	paradise, opinion	crouton, guitar, member, pidgin, psychologist, tantrum
CONCAVE	putsch, chattel	rift, din, dupe, hake, ling, tuque
BUSH	dearth, tallow	dirge, dolt, dowel, dower, loll rev

Table 2: Targets and Fillers: Four Neighbourhood Types

To select words to be used as stimuli, this study utilized Hoosier Mental Lexicon (HML) in Nusbaum et al. (1984) and its familiarity rating values. Before setting the range of familiarity for each type, its distribution in the HML was investigated. It showed that the familiarity rating on words were heavily skewed, and the mean familiarity rating was 5.88 in a 1-to-7 scale. Thus, the present study defined highly familiar words as words from the third quantile to the maximum of familiarity rating; for low familiar words, it ranged from the minimum up to the first quantile (min=1.000 Q1=5.083, Q3=7.000, max=7.000). In addition to the specification of target familiarity values, the neighbourhood density of each lexical item and the average familiarity value of neighbours were calculated (min=1.000, Q1=5.625, mean=5.980, Q3=6.750, max=7.000). Then, the words which had a larger difference between target familiarity and mean neighbourhood familiarity were chosen as POLE or CONCAVE. For FOREST and BUSH, the words with a smaller familiarity difference between target and neighbours were chosen. As for the structure of words, the word length by syllable count of each neighbourhood type was controlled. Most of the stimuli were three syllable words, except for the FOREST type, due to the limited availability of words with the required familiarity specifications. Proper nouns, loan words, and polysemies were excluded from the word list. Lastly, each selected word was pronounced by a female native English speaker in her 20s from Australia, and was digitally recorded for audio stimuli.

The present study assumed that learners also show the performance influenced by lexical effects. The hypotheses of the present study are as follows: the learners of English will show familiarity effect in a retrieval task; high familiarity words will be responded faster than low familiarity words. Likewise in Vitevitch and Rodriguez (2005), neighbourhood effect will appear in high-familiarity conditions (FOREST and POLE). In addition to the familiarity effect and neighbourhood effect, learners at higher level of proficiency will respond faster than those at lower level in high-familiarity conditions (FOREST and POLE), assuming that more fluent learners will show similar response patterns to those of natives. To summarize, the present study predicts there will be a familiarity effect and a neighbourhood effect on learners' lexical processing, with an interaction of learners' proficiency.

2.3 *Procedure* The participants performed the experiment in either a soundproof room or a quiet classroom in the universities where they were from. The participants used a Macintosh computer with PsychoPy software version 1.85.6 (Peirce, 2015; see also Peirce 2007, 2009). They put a set of headphones (Pioneer SE-M531) on and listened to the stimuli at their comfortable level. The experiment mainly consisted of two tasks: memorization and retrieval. The participants were asked to memorize eight target words by listening to each sound file in Task 1. They could repeat the target sounds as many times as they want for memorization. In Task 2, they listened to the targets and the fillers one by one and judge if they heard it in the memorization task or not, using designated keys (J for the targets, and F for words they did not memorize). Participants listened to 32 stimuli, which contained 8 targets and 24 fillers repeated twice and randomized, and they completed entire Task 2 twice. Thus, each participant heard 128 tokens in total. The flow of stimuli presentation and measurement is as follows: first, the mask sign (+) was displayed for 1.2 seconds, then the auditory stimulus was presented via the headphones. The participants were instructed to press the key for judgment as accurate and fast as they can. About four seconds later, the mask sign was presented on the display as an indicator for the upcoming sound. Figure 3 illustrates the flow of Task 2, with its numbers in seconds.

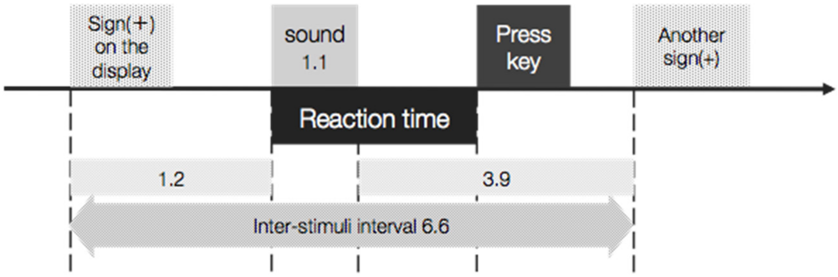

Figure 3: The Flow of Retrieval Task (Task 2, in seconds)

Reaction time for each stimulus was obtained by measuring the duration between the onset of the sound and the onset of pressing the keys. If participants press keys other than J and F, their responses were not included in the reaction time calculation. If they press the correct key more than once, the first time stamp was used for the calculation.

3 Results

3.1 *Reaction Time and Neighbourhood Types* The participants' performances were analysed by the linear mixed model on R [3.5.1] (with lmerTest [3.0-1] package). First analysis used their reaction time data as a dependent variable and neighbourhood types with two random effects: subjects and words. Figure 4 shows the mean reaction time to each word type. The order of each bar is BUSH, CONCAVE, FOREST, and POLE from left to right. The reaction times to FOREST and POLE types were significantly shorter than those to BUSH (Forest t=-2.529 [*df*=36.89223], *p*<.05; Pole *t*=-3.063 [*df*=36.86675], *p*<.01).

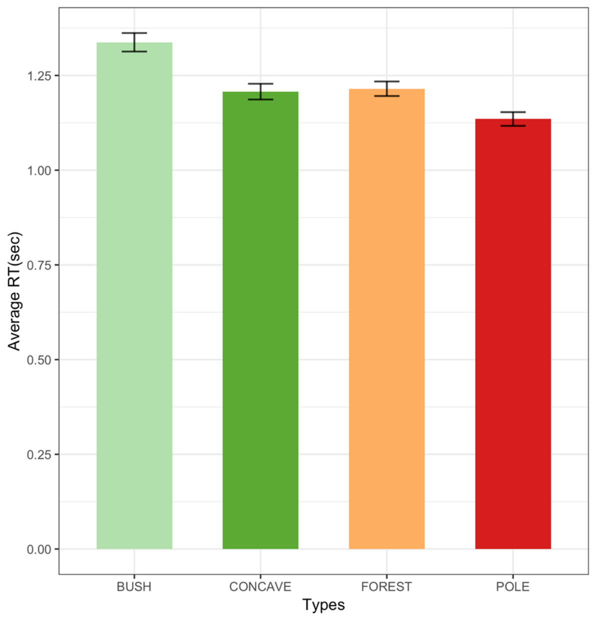

Figure 4: Mean Reaction Time of Four Neighbourhood Types

Another analysis aimed to confirm proficiency effect, and this analysis used both proficiency difference and neighbourhood types as independent variables with the same random effects as in the previous analysis. It showed that there was an interaction between English proficiency levels and the FOREST type. Advanced learners responded significantly faster than intermediate learners in FOREST condition (*t*=-0.10741 [*df*=3001.04560], *p*<.05). No other interaction was significant. Still, the main effect of proficiency was confirmed; the mean reaction time of the intermediate level was 1.345962 sec, which is significantly longer than advanced learners' average (Advanced level Mean RT= 1.101450; *t*=0.23201 [*df*=25.78365], *p*<.05). Figure 5 illustrates the mean reaction time of each neighbourhood type for advanced learners and intermediate learners. The dark bar on the left side in each pair represents mean reaction time of advanced learners; the other bar represents that of intermediate learners.

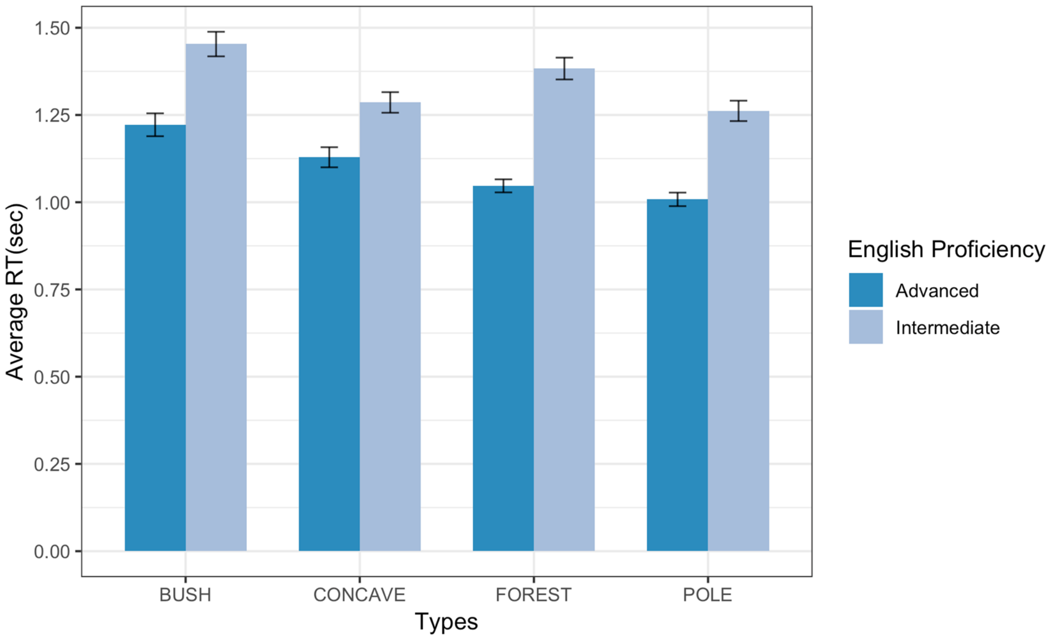

Figure 5: Mean Reaction Time of Advanced and Intermediate Learners

3.2 *Reaction Time and Lexical Properties* Instead of reaction time analysis based on four neighbourhood types, linear mixed model with actual values of word familiarity, mean neighbourhood familiarity, and neighbourhood density as independent variables was conducted to clarify confounding effect of lexical properties. Several models differing in the number of neighbourhood-related variables were created and compared to test the significance of neighbourhood effect as a function to reaction time. Since four neighbourhood types took both target and neighbourhood value together, the previous analysis could not detect if the lexical effect derives from two properties or only one of them. Table 3 summarizes the comparison of models differing in the number of independent variables.

Models	*AIC*	*Pr(>Chisq)*
A: rt ~ Fam +(1| subj)	4617.6	-
B: rt ~ Fam*mFamN +(1| subj)	4581.2	1.694e-09 ***
C: rt ~ Fam*mFamN*ND +(1| subj)	4536.0	7.562e-11 ***

Table 3: Post hoc Analyses on Neighbourhood Effect (Fam: target familiarity, mFamN: mean neighbourhood familiarity, ND: neighbourhood density)

Model B and Model C, which included variables of both targets and neighbours, were significantly better than Model A, which had the target familiarity only. This test supported the idea that both target and neighbourhood lexical properties have a significant effect on reaction times; those combinations of lexical properties were incorporated into the four categorizations. Further discussion on comparing type-based model and a continuous model will be given in section 4.

3.3 *d-prime* To confirm the participants' accuracy in the retrieval task, the d-prime (d') score for each participant was calculated by using *dprime.SD* function in *psyphy* package [0.1-9] (Knoblauch, 2015) on R. The d-prime value was calculated by the difference between z-transformed hit rate and false alarm rate (for the details of d-prime and its theoretical background, see Macmillan and Creelman, 1991). In the present study, hit means participants' correct key press for the target stimuli; false alarm means the participants' incorrect key press for the fillers, i.e. responding “Yes” to non-memorized words (fillers). The d-prime score for each participant and English proficiency variable, a classification within the CEFR scheme, were submitted to a linear regression analysis. Figure 6 shows the distribution of d-prime values for participants in each CEFR category. According the the linear regression analysis, all levels of participants were significantly different; the participants at the higher level were more accurate to the stimuli than those of lower (C1 t=21.133, p<.01; B2 t=28.956, p<.01; B1 t=3.622, p<.01). The d-prime difference still remained significant in a rather coarse advanced-intermediate grouping as in the reaction time analysis. Advanced learners had significantly higher d-prime value than those of intermediate learners at .001 significance level. The difference between intermediate and advanced was 1.15680 (t=-32.74 [df =3070], p<.01). The statistical test showed that the sensitivity to the stimuli in L2 correlated with the proficiency level of the participant.

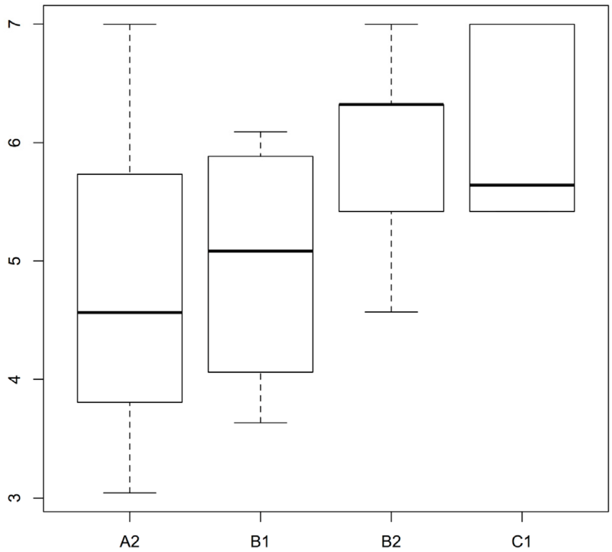

Figure 6: d-prime Score according to CEFR Categorization

4 Discussion and Conclusion

From the two analyses of reaction time in 3.1, the present study confirmed the lexical effect on L2 learners’ speech perception. The first comparison of reaction times among neighbourhood types based on lexical properties confirmed a positive familiarity effect even for learners, in that the reaction time of highly familiar targets (FOREST and POLE) were significantly shorter than those of BUSH. The present study confirmed the facilitative effect of familiarity as in Connine et al.’s study (1990). The other statistical analysis revealed that the interaction of English proficiency and lexical properties was limited to the FOREST type, which has both highly familiar target and highly familiar neighbours. These two results confirmed Vitevitch et al. (2005)’s study in terms of word frequency, but based on a different lexical property: familiarity. Although both Vitevitch et al. (2005) and Connine et al. (1990) are about L1 speech perception, the present study revealed that familiarity and neighbourhood effect is even valid for L2 speakers. In addition to reaction time analysis based on neighbourhood types, post hoc analyses in section 3.2 proved that the neighbourhood effect is a significant factor to account for the reaction time variance.

As well as lexical effect of familiarity and neighbourhood for L2 learners, the present study investigated the learners’ proficiency as a key factor to their performance. The significant difference between advanced and intermediate learners is the ability to judge among the highly familiar words in the perception and recognition process of speech. Given that highly familiar targets were recognized faster than low familiarity targets in both levels of proficiency, and that the interaction between English proficiency and word properties was present, we can infer that learners improve their speech processing in a selective way. That is, the learners set a priority to highly familiar words over less familiar words in speech perception. The absence of interaction between English proficiency and POLE type also supports the idea of selective development of learners’ lexicon. Since the neighbours of POLE type were not familiar to all subjects, POLE type had no significant neighbourhood effect to the learners at both levels of proficiency. On the contrary, FOREST had an interaction between language proficiency and neighbourhood, for FOREST had both highly familiar targets with highly familiar neighbours. Therefore, the nature of retrieval task in POLE and FOREST conditions actually revealed the learners’ different levels of performance. FOREST condition tested the subject’s ability to detect the target by rejecting highly familiar candidates surrounding the target, whereas POLE condition tested their ability to obtain the stored word from their working memory, possibly having little to do with linguistic experience. In addition to the reaction time difference among learners, the d-prime comparison confirmed the development of learners' sensitivity to the stimuli. Advanced learners scored significantly higher d-prime value than intermediates; learners improve their accuracy in retrieving their memory better by taking advantages of their lexical knowledge. That is a viable explanation why the advanced learners were better at judging only in FOREST condition than intermediate learners, but not in POLE condition.

The limited reaction time difference in relation to learners' proficiency may be attributed to L2 learning process. Considering the fact that the L2 learners' inputs may contain more familiar words than less familiar ones due to the limitation of L2 vocabulary, they may enhance specialized connections between a lexical item and the more familiar string of sounds in L2 lexicon. Thus, as they get more inputs and strengthen the connections in their lexicon, such development may ameliorate both speed and accuracy of determining the target word and deleting neighbouring candidates.

The present study gave an insight to L2 learners' speech perception, in relation to the previous L1 studies. Although it revealed that there are familiarity effect and neighbourhood effect on L2 learners, the subjective evaluation of word familiarity by L1 speakers may not be completely identical to L2 lexicon. The nature of lexical familiarity is still unclear as well. However, this study revealed that familiarity judged by natives could be used as a possible benchmark for L2 speakers' mental lexicon, in that our results were at least consistent with the results of L1 studies. To compare learners' mental lexicon to that of natives overtly, the future study must need to collect L2 learners' subjective familiarity to the stimuli.

NAM with four distinct types (corresponding to neighbourhood types used in our study) in Luce and Pisoni (1998) did not assume any specific mapping of lexical items, which would be continuous or categorically skewed. Therefore, the four neighbourhood types could be just extreme cases of the distribution of the target frequency and the mean neighbourhood frequency. The lexicon could have many possible states in between, and the mapping could be continuous if those in-between states were all filled up. Based on such consideration on the mapping in lexicon, a post hoc test comparing a type-based model and a continuous variable model brings new discussion questions in relation to L2 proficiency effect. Two models differing in independent variables, one of which had categorical variable (Model I) as used in section 3.1 and the other had continuous variables of familiarity and mean neighbourhood familiarity (Model II), were created, and an ANOVA was conducted to compare them. This analysis was aimed to see if there was any difference between two assumptions of learners' lexicon. Although there was no significant difference between these two models, Model I had slightly smaller AIC than Model II (I: AIC=3674.7; II: AIC=3676.3). Interestingly, Model I showed a significant proficiency effect as an interaction with FOREST type, whereas proficiency difference in Model II had no significant effect on reaction time. There are two possible accounts for the result of this post hoc analysis. First account derives from the participants' proficiency divergence, that is, the divergence of participants' proficiency was not wide enough to be a significant factor to the task. Another account comes from the nature of leaners' lexicon, although it is still obscure. The present analysis showed that the categorical model was slightly better than the continuous model, and it incorporated with proficiency effect as one of significant factors. In addition, participants' d-prime scores were significantly different from each other. Given the sensitivity difference among various proficiency levels, the better fit in Model I may emerge from the interaction between learners' proficiency and lexical effects. As discussed in the second paragraph of this section, the advanced learners improve their perception selectively, giving a priority to high familiarity words over low familiarity words. This selective development of lexicon may also contribute to the better fit of Model I than Model II. In other words, the Model I may reflect the nature of learners' lexicon better than Model II. Since the present study did not originally aim to compare categorical and continuous models of the learners' lexicon, this argument is just a speculation based on the post hoc analysis. The geographical structure of mental lexicon in L2 still remains largely unexplored, and its relation to L2 learning process may be the possible area of future study.

References

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(6), 1084.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. Journal of experimental psychology: General, 113(2), 256.

Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. Word, 20(2), 157-177.

Knoblauch, K. (2015). Functions for analyzing psychophysical data in R. [Package for R]. Retrieved from https://cran.r-project.org/web/packages/psyphy/psyphy.pdf

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 Memory and Language, 12(2), 119.
Luce, R.D. (1959) Individual Choice Behavior: A Theoretical Analysis. Wiley, New York.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. Ear and Hearing, 19(1), 1–36.
Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guideCambridge University Press. New York.
Ministry of Education, Culture, Sports, Science and Technology - Japan (MEXT). (2018).Kaku-shiken, kentei kekka to CEFR tono taisyou-hyou. [Correlation Table of English Standardized Scores and CEFR]. [PDF file] Retrieved from http://www.mext.go.jp/b_menu/houdou/30/03/__icsFiles/afieldfile/2018/03/26/1402610.pdf
Morton, J. (1969). The interaction of information in word recognition. Psychological Review, 76, 165-178.
Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, S. Franck, E. C. T. Walker, & M. Garrett (Eds.), Perspectives on mental representations, (pp. 89- 109). Hillsdale, NJ: Erlbaum.
Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon. Research on spoken language processing report no, 10, 357-76.
Peirce, J.W., (2007). PsychoPy - Psychophysics software in Python. J Neurosci Methods, 162(1-2):8-13
Peirce, J.W., (2009). Generating stimuli for neuroscience using PsychoPy. Front. Neuroinform. 2:10. doi:10.3389/neuro.11.010.2008
Peirce, J. (2015). PsychoPy version 1.85.6. [Computer Software]
Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. Speech communication, 4(1-3), 75.
Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1R17
Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. Journal of memory and language, 39(3), 458-483.
Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. Journal of Multilingual Communication Disorders, 3(1), 64-73.