



Eva Maria Luef

Charles University in Prague, The Czech Republic

Pia Resnik

University College of Teacher Education Vienna/Krems, Austria

 <https://orcid.org/0000-0002-2362-2422>

 <https://orcid.org/0000-0003-0948-9546>

Phonotactic Probabilities and Sub-syllabic Segmentation in Language Learning

Abstract

High phonotactic probabilities are known to exert a facilitative effect on word learning in children and adults in their first language. The present study was designed to investigate the role of phonotactic probabilities when learning a foreign language. Focusing on Austrian and Korean learners of English, we investigated two hypotheses related to phonotactic frequency effects: (1) High-frequency segments have more deeply entrenched phonetic representations, with more automatized pronunciation patterns, rendering phonetic learning of homophonous segments more difficult; (2) High-frequency segments are associated with higher phonetic variability in the first language, which can facilitate phonetic learning in a foreign language. Additionally, the locus of phoneme/ bigram frequency effects was analyzed in relation to left-branching and right-branching syllable structure in German and Korean. We found that proximity to English voice-onset time is correlated with phoneme and bigram frequencies in the first language, but results varied by learner group. Sub-syllabic segmentation of the first language was also shown to be an influential factor. Our study is grounded in research on frequency effects and combines its central premise with phonetic learning in a foreign language. The results show a tight relationship between first language statistical probabilities and phonetic learning in a foreign language.

Keywords: Austrian German, English as a Foreign Language (EFL), frequency distribution, Korean, sub-syllabic segmentation

Background

Phonotactic probability is defined as the position-specific frequency of segments and segment combinations (Vitevitch, 1997; Vitevitch & Sommers, 2003) and is thus a measure of how frequent (and probable) particular segments of words and sequences of phonemes are (Vitevitch & Luce, 1999). Different phonotactic constraints apply to different languages and (first and foreign) language learners accumulate knowledge on phonotactic probabilities based on experience (Weber & Cutler, 2006). High-frequency phonotactic combinations serve an important purpose in word recognition, as words including such combinations are generally recalled faster and more accurately (Frisch, Large, & Pisoni, 2000; Luce & Large, 2001; Vitevitch, Armbruster, & Chu, 2004; Vitevitch & Luce, 1998). High phonotactic probability has not only been linked to more rapid word learning in adults but also in child language acquisition (Storkel, 2001; Storkel & Maekawa, 2005; Storkel & Rogers, 2000). The advantage in word learning involving high-probability phonotactic combinations could result from strengthened cognitive representations of the frequent phonotactic combinations (Bybee, 2007). Storkel (2001), for example, suggested that high phonotactic probability segments also influence the formation of semantic representations and the association between semantic and lexical ones, thus furthering learning.

While some studies have linked phonotactic probabilities to word learning in general (e.g., Storkel, Armbruster, & Hogan, 2006), less is known about phonetic learning. Based on previous work on word frequencies, several predictions can be inferred regarding frequency and probability effects in relation to phonotactic combinations. It has been shown that high-frequency words may be more deeply engrained in linguistic memory, and thus have more entrenched phonetic patterns (Bybee, 2007; Levy & Hanulikova, 2019; Pierrehumbert, 2001; Schweitzer et al., 2015). The special role of high-frequency distributions of particular words in connection to phonological changes has long been acknowledged in studies on linguistic change (Bybee, 2002; Phillips, 1984; Pierrehumbert, 2001). Under certain circumstances low-frequency words may be phonetically more malleable and thus more prone to sound change than high-frequency words (Phillips, 1984, 2006; Todd, Pierrehumbert, & Hay, 2019). An alternative hypothesis describes high-frequency words as having larger exemplar clouds, that is, being associated with more phonetic variation in the speaker's mind (Levy & Hanulikova, 2019; Schweitzer et al., 2015). This implies that speakers have more numerous and diverse phonetic targets associated with each high-frequency speech sound. Low-frequency sounds have smaller exemplar clouds and thus show less phonetic variability (Levy & Hanulikova, 2019). The crucial difference between these hypotheses is whether high-frequency rates limit or increase variability, and this has implications

not only for sound change but also for language learning. While the above-mentioned studies focused on the word level, similar tendencies may be at work at the segmental level. Lexical frequency rates and phonotactic probabilities have shown to be correlated (Storkel & Maekawa, 2005) in English, allowing the cross-fertilization of theories in the two strands of linguistic investigation.

What is suggested for phonological change may also apply to foreign language learning of novel phonetic detail in a known phonotactic combination (i.e., cross-linguistic phonotactics). When learners of a foreign language encounter a high-frequency phonotactic combination that is similar to one in their first language (e.g., /bi/), they may either be phonetically limited by their first language, or they may have access to a highly variable phonetic inventory and thus be better able to approximate the foreign-language phonetics. In contrast, low-frequency phonotactic combinations in the first language may be either more malleable to phonetic learning due to their shallow cognitive entrenchment, or learners may have a smaller phonetic inventory and face more difficulty in finding a suitable pronunciation. The two hypotheses lead to very different predictions with respect to how learners can acquire the phonetics of phonotactic combinations in the foreign language. The following study focuses on learners of English as a Foreign Language (EFL) and investigates how phonotactic probabilities of utterance-initial segments in their first languages (Korean, German) impact phonetic learning of the cross-linguistic variants of the combinations in English.

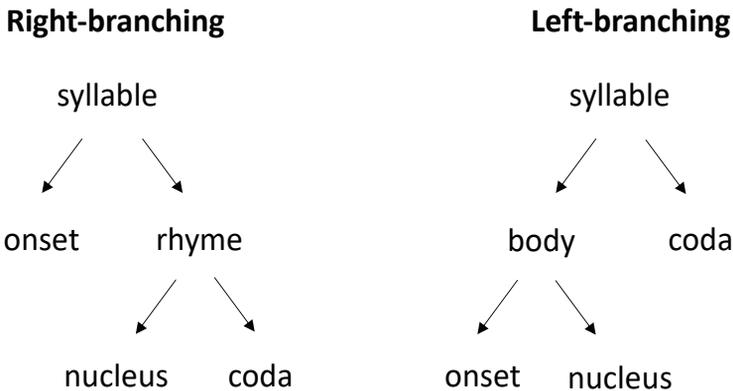
Korean and German are typologically different languages, and one key difference concerns the structure of the syllable. While syllable universals have been hard to define, the general outline of onset-rhyme (i.e., right-branching syllables) and body-coda (i.e., left-branching syllables) is an accepted categorization (Berg & Koops, 2010; J.-Y. Kim & Lee, 2011). The difference between the two types is the linkage strength between the initial segments. Whereas the onset-rhyme structure separates the initial phoneme from the rhyme in closed syllables, the body-coda system binds the initial phoneme and the following vowel together (J. Kim, 2015). For instance, a syllable such as /ban/ would be perceived with /b/ separate from /an/ in the German onset-rhyme structure, whereas in the Korean body-coda structure, /ba/ would go together and /n/ would be perceived as a separate entity (see Figure 1).

Berg and Koops (2010) and Kim (2015) speculate whether the left- and right-branching preferences found across Korean and English are also related to phonotactic dependencies between segments. How robustly the nucleus vowel is formed in phonetic memory in connection with either the onset or the coda is unclear at the moment. Phonotactic probabilities have been shown to have an effect on the perception and processing of syllable structure, with Korean speakers being better at processing the onset and nucleus of a syllable rather than only the initial phoneme (J. Kim, 2015; J. Kim & Davis, 2002; Witzel, Witzel,

& Choi, 2013). The sub-syllabic characteristics of Korean indicate that initial bigrams are a crucial unit in speech processing in the language. In German, the initial segment may be more influential.

Figure 1

Sub-syllabic Structuring in Right-branching and Left-branching Syllables
(J.-Y. Kim & Lee, 2011)



The present study analyzes phonotactic probabilities of word-initial phonemes and bigrams (or biphones) in English, Korean, and German, and relates them to phonetic learning of English as a Foreign Language in speakers of Korean and German. The following two inter-linked research questions are posed:

1. Are high-frequency phonotactic combinations more difficult to adapt through learning than low-frequency phonotactic combinations?
2. Does sub-syllabic structure play a role? Specifically, is Koreans' EFL speech more strongly impacted by initial bigram frequencies, while Germans' EFL speech is more strongly influenced by initial phoneme frequencies?

Two groups of EFL learners, Korean first language (L1) users from Seoul and Austrian speakers of L1 German, are compared in terms of phonetic learning of voice onset time in word-initial fortis and lenis plosives in English. Confounding factors that may influence phonotactic probability and/ or word-initial voice onset time (VOT), such as lexical frequency rates, neighborhood density, English phoneme and bigram frequency, and EFL phoneme and bigram frequency are considered in the analysis.

Voice onset Time in English, Korean and Austrian Plosives

English distinguishes two phonation types of plosives, commonly referred to as “lenis” and “fortis” (or “voiced” and “voiceless”). In utterance-initial position, American English lenis plosives are phonologically voiced, phonetically voiceless and unaspirated, with a mean VOT range of 8 to 17 msec. (Chodroff, Godfrey, Khudanpur, & Wilson, 2015). The utterance-initial fortis plosives are phonologically and phonetically voiceless and aspirated, with a mean VOT range of 65 to 120 msec. in American English speakers (Berry & Moyle, 2011). In other positions, including word-initial but utterance-medial, American English plosives are more likely to have voicing (Davidson, 2016). Regional differences in VOT have been noted, with speakers from Southern states displaying a tendency to pre-voice word-initial lenis plosives (Hunnicuttt & Morris, 2016; Morris, 2018). Lenis VOTs of speakers from Southern British English (e.g., London) range from 10–22 msec. (Sonderegger, 2015), but speakers from Scotland may show significant pre-voicing of up to 100 msec. (Watt & Yurkova, 2007). British English fortis VOTs most frequently range between between 50 and 100 msec. for Northern England and Scottish speakers (Docherty, Watt, Llamas, Hall, & Nycz, 2011) but are shorter for Southern England speakers, ranging between 35–75 msec. (Sonderegger, 2015). There is significant overlap between British and American English voice onset times, and both are different from Austrian German and Korean in certain respects.

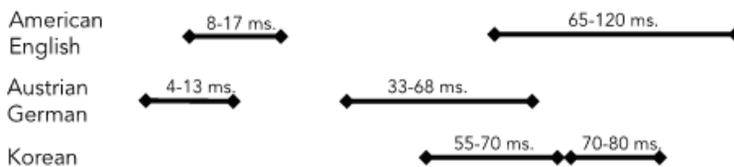
The terms lenis and fortis are also used to describe the two phonation types of German plosives. German shows no voicing of plosives in word-initial position and has longer VOTs of lenis plosives than English. Southern German (including Austrian) plosives differ from Northern/ Middle German and a near-merger of word-initial fortis-lenis contrasts in some articulatory positions complicates the pattern (Moosmüller & Ringen, 2004; Moosmüller, Schmid, & Brandstätter, 2015). Aspiration is absent (Moosmüller, 1987; Siebs, de Boor, Moser, & Winkler, 1969) and contemporary Austrian lenis plosives are characterized by short VOTs, while fortis plosives show no aspiration in bilabial, little aspiration in alveolar and strong aspiration in velar position (Luef, 2020). In younger Austrian speakers, who are in the process of phonetically splitting the near-merger, mean lenis VOTs range between 4 and 13 msec., while fortis plosives show an average range of 33 to 68 msec.

Korean shows a three-way distinction in plosives (lax or lenis, aspirated, and tense, see J. Y. Kim, 2010; Shin, Kiaer, & Cha, 2013). The so-called fortis plosives in Korean usually refer to the tense category (e.g., [p*], characterized by very short VOT), and are thus not equivalent to the Germanic fortis plosives. The Korean lenis plosives are phonologically voiceless and show mean VOT values of approximately 55 to 70 msec.; the aspirated plosives are phonologi-

cally voiceless, with long VOT values between 70 and 80 msec. (Kang, 2014; Silva, 2004, 2006). A merger of lax and aspirated plosives in all three articulatory positions has led to VOT overlap of phrase-initial lenis and aspirated plosives (Jucker & Smith, 2006; Silva, 2006). Recent studies have shown that the VOT ranges for lax stops have increased, while those for aspirated ones have decreased, with the VOT difference between these two categories reducing accordingly (Chang & Kwon, 2020). This change in Korean has implications for the realignment of the Korean and English stop categories, with both lax and aspirated stops approximating the VOT ranges associated with English voiceless stops, as schematized in Figure 2. While the Korean plosive merger has obscured phonetic distinctions between lax and aspirated plosives, the F0 distinction at the onset of the following vowel has been amplified: the F0 values for aspirated stops are higher than those of lax stops, a trend that has led to distinct tonal levels (Kang, 2014). The vowel environment of a word-initial plosive can have influences on VOT duration in different languages (Esposito, 2002; Grassegger, 1996; Klatt, 1975; Moosmüller & Ringen, 2004; Mortensen & Tøndering, 2013). Vowel height plays a role here and constricting the air passage through the vocal tract (such as when raising the tongue) will lead to a delay in voice onset time (Fischer-Jørgensen, 1980). Thus, high vowels will cause VOT to be prolonged, while low vowels cause it to be shortened.

Figure 2

Mean VOT Values of Short-lag VOT (Lenis, Lax) and Long-lag VOT (Fortis, Aspirated) in American English (Based on Berry & Moyle, 2011, and Chodroff et al., 2015), Austrian German (Based on Luef, 2020), and Korean (Based on Kang, 2014 and Silva, 2004, 2006)



The mapping of Austrian and Korean plosives onto English ones is phonetically complicated. Austrian lenis and American English lenis can be regarded as corresponding; however, Austrian fortis only has small overlaps with American English fortis. The Korean lenis category ranges within the Austrian fortis category, with significant overlaps with American English fortis

plosives. Korean aspirated plosives range within the American English fortis plosives. While phonetic mapping of the three languages is difficult, grapheme mapping is clear. German and English graphemes of lenis and fortis plosives are identical and German readers of English will immediately map them correspondingly. A widely used language Romanization system in South Korea (“Revised Romanization of Korean”/ 국어의 로마자 표기법) transcribes the word-initial lenis plosives <ㅂ>, <ㄷ>, and <ㄱ> as , <d>, and <g> and the aspirated plosives <ㅃ>, <ㄸ>, and <ㄲ> as <p>, <t>, and <k> (note: tense plosives are transcribed with double consonants, e.g., <bb>). Here, grapheme correspondences between Korean and American English lenis and aspirated/fortis categories are established and may guide Korean readers of English in their mapping of plosive correspondences. The present study tests phonetic learning of Korean and Austrian learners of English and is based on reading stimuli. Therefore, grapheme mapping is expected to be influential in the process. Austrian learners certainly map their lenis and fortis contrasts onto the English lenis/fortis distinction, and Korean learners may be more inclined to map their lenis onto the English lenis and their aspirated contrasts onto the English fortis category.

According to the *UCLA Phonological Segment Inventory Database* (see Maddieson, 1984), plosive consonants (especially fortis) are among the most frequent phonemes in languages world-wide (also see Everett, 2018). Even though individual languages utilize them to different degrees, their articulatory and perceptual ease makes them pervasive to the human language capacity (Ohala, 1983). From such a universal view of phonological complexity (e.g., Romani, Galuzzi, Guariglia, & Goslin, 2017), it could be assumed that differences between their individual frequency rates may not lead to significant differences in foreign language learning.¹ In usage-based accounts of language acquisition and development, phonemic frequency generally plays a role, with different predictions resulting for production and perception of phonemes (Bybee, 2001). Studies have shown that VOT contributes to transfer effects in second language learners (e.g., Schoonmaker-Gates, 2015; Skarnitzl & Rumlová, 2019), suggesting an effect of language-specific phonological patterns, which impede or facilitate phonological learning in a second language.

¹ We are grateful to an anonymous reviewer for pointing this out.

Methods

Participants and Procedures

Speakers whose first language was Korean ($N = 22$, male: 5; female: 17) and Austrian German ($N = 21$, male: 3; female: 18) were recruited in their home countries in the cities Seoul and Vienna, respectively, for a sentence-reading task in their foreign language English. Participants were students whose ages ranged from 19 to 27 (mean = 23.2), and who were enrolled in foreign-language programs at their respective universities (Seoul National University, University of Vienna), where admission required English proficiency levels of B2 or higher according to the Common European Framework of Reference for Languages (Council of Europe, 2018). The majority of students were in advanced years of their program, some of them in graduate programs. They primarily reported using their first languages in their daily lives but were highly exposed to American English through online media and, in the case of Koreans, by American pronunciation teachers (Ahn, 2011). Austrian students of English may be exposed to British English to a higher degree, having travelled to Great Britain or being tutored by British pronunciation teachers. All participants were first informed about the recording procedures (but not told about the objective of the study) and their rights as participants. After having given their consent, they completed a survey that collected demographic information and details about the participants' linguistic habits (e.g., first language, dialect, exclusion of speech impediments). The participants were paid for their participation and the experiment took place between November 2018 and June 2019. The study compared two experimental groups but no control group was included in the experimental design.

The sentence-reading task consisted of 86 short English sentences or phrases (mean words per sentence = 6.3, $SD = 2.1$) which were read once at a comfortable speed and in the same order by each participant. The sentences were typed with a word processor and printed on a piece of paper that was given to each participant. Each sentence contained a target lexeme with a word-initial plosive consonant in sentence-initial position (e.g., 'Buffaloes are large animals' or 'Cats are active at night,' see supplementary material Table A1 for the list of carrier sentences), resulting in similar prosodic/rhythmic structure of the sentences. Participants were not familiar with the sentences and phrases before the start of their reading and were asked to assess the level of difficulty afterwards in their first language by speaking aloud the terms for 'easy,' 'medium,' and 'difficult' (Sino-Korean: 'ha': 하, 'jung': 중, 'sang': 상; German: 'leicht,'

‘mittel,’ ‘schwer’). By uttering a Korean or German term after each sentence/phrase, we attempted to minimize habituation effects. The order of word-initial plosive appearance was shuffled so that no consecutive sentences started with the same plosive. All target lexemes had the primary stress on the first syllable. Each plosive type (lenis/lax and fortis/aspirated variants of bilabials, alveolars, and velars) appeared in word-initial bigrams with high vowels ([i, ɪ]), mid ([e, ε, æ]), and low vowels ([a, ʌ, ɔ]). We grouped the vowels according to height in order to account for the VOT differences in relation to vowel height. Each bigram combination appeared a minimum of four times, resulting in each plosive type appearing at least 14 times throughout the sentence-reading task. For instance, the bigram [di, di] started the five sentences ‘Deans of colleges have to work long hours,’ ‘Dishwashers are too expensive for me,’ ‘Deals in the business world are hard to make,’ ‘Differences in opinion should not be expressed,’ and ‘Dill is an herb used for Italian cooking.’ Sentences belonging to the same bigram class (e.g., lenis alveolar + i/i) were spaced apart at a minimum of ten sentences. All sentences were semantically unrelated to their neighboring sentences and no phonological neighbors in target words were presented in consecutive sentences. Cases of deviant phonology (e.g., [dʒɪl] instead of [gɪl]) or stress placement (e.g., ‘dessert’ instead of ‘desert’) were removed from the sample. The possible difference in isochronous temporal patterns between Korean (Lee, Jin, Seong, Jung, & Lee, 1994) and German (Port, 1983) was negligible in the present study as only sentence-initial syllables with primary stress were the focus of analysis.

The participants’ speech was recorded with a ZoomH4n digital audio recorder with an attached Sennheiser ME67 microphone. Speech was sampled at 44.1 kHz at 16-bit depth, and was subsequently saved and stored as .WAV files. Target lexemes were cut manually from the audio stream and saved as separate files, which were later processed with the open-source acoustic software Praat (Boersma & Weenink, 2019). Overall lexeme duration as well as the duration of the word-initial VOTs were manually annotated on two different tiers in the program that allowed automated extraction of the durations (in seconds) via a script.

The start of each lexeme/VOT was marked at the burst of the stop (Abramson & Whalen, 2017); the end of VOT was determined at the onset of glottal pulsing (settings: 100-600 Hertz for women and 75 to 300 Hertz for men, Vogel, Maruff, Snyder, & Mundt, 2009). The majority of words (78%) ended in alveolar fricatives (of which 97% were <s>, voiced or unvoiced) and here the end point was marked when the frication had ceased (i.e., the nearest zero crossing) as visible on the waveform and spectrogram. In the case of plosives (10%), nasals (7%), liquids (3%), or vowels (2%) constituting the final phonemes of the target lexemes, the end point was determined when the waveform cycle had ceased and the sound had completely faded.

VOT was normalized for speech rate by calculating a measure of syllables per second on 5% of each participant's speech (= eight sentences per participant taken from the middle of the reading texts; the sentences were the same for each participant). This value was then multiplied with VOT (in seconds) and later converted to milli-seconds by multiplying it by 1000.

Approximately 7% of the data was coded for reliability by a second observer and Pearson's *R* along with the root mean square error (RMSE) were calculated to see whether the two coders agreed on (a) overall word duration and (b) start of VOT (= initial burst). For word durations, an excellent *R* value of .99 (RMSE = .023) and for VOT durations, an acceptable *R* value of .71 (RMSE = .014) are reported.

Variables *VOT Distance*

In order to determine the degree of similarity of the Korean and Austrian learners' VOTs to those of native English speakers, VOTs of American English speakers were extracted from the TIMIT Corpus, a collection of sentences read by American English speakers from different dialect regions, which is widely used in the phonetic sciences (Garofolo et al., 1993). Even though American English VOTs show socio-phonetic and regional stratification (see, e.g., Lipani, 2019), the present study will focus on average VOTs across the variety of American English speakers. We identified sentences starting with nouns with initial bigrams that were the focus of our study (see Participants and Procedures). Primary stress had to be on the first syllable ($N = 146$). We measured VOT in the identical way as described for the EFL learners. Due to an underrepresentation of the sentence-initial bigrams b, d, and g plus [i, ɪ], d and g plus [e, ε, æ], selected recordings of the American radio show "This American Life" (<https://www.thisamericanlife.org>) were added to the corpus ($N = 21$). After identifying speakers whose biographical information (e.g., age) were available, bigrams representing the initial segments of sentence-initial nouns were manually cut from the .WAV files that were downloaded from the website of the show. Acoustical measurements followed the procedures as outlined for the EFL learners and the TIMIT Corpus. Speech rates of each American English sentence in the TIMIT corpus were calculated (syllables per second) and each VOT was normalized for speech rate. See Appendix Table A2 for more information on the American English speaker data.

The phonetic distance between the Korean/Austrian VOTs to the American English target VOT spaces was assessed by calculating the Mahalanobis distance (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015), which computes the distance of a test point from the distribution mean by considering the covariance matrix (Martos, Muñoz, & González, 2013). The Mahalanobis

distance takes into account natural variability in speech production by calculating the number of standard deviations from a learner's VOT to the mean of the target spaces (computed per plosive type) derived from the American English speakers, along each principal component axis of the target spaces (Kartushina et al., 2015). A Mahalanobis distance of 0 indicates that a learner VOT value is at the mean of the target space. After analyzing z-scores of Mahalanobis distance scores and removing those over three standard deviations, the highest Mahalanobis distance in the present study was 14.21.

Frequency Variables

Frequency rates of Korean initial phonemes were taken from Shin, Kiaer, and Cha (2013) who based their calculations on the *Yonsei Korean Language Dictionary* and the *Standard Korean Language Dictionary* in combination with the SLILC Spoken Language Information Lab Corpus (Shin, 2008). To determine the frequency rate of plosive-plus-vowel bigrams in word-initial position in Korean (which are not included in Shin et al., 2013), we used the Korean corpus of the *Leipzig Corpora Collection/ Deutscher Wortschatz Corpus*, comprising over 109 million tokens and over seven million types extracted from Korean newspapers between 2011 and 2019 (Goldhahn, Eckart, & Quasthoff, 2012). We analyzed the first 100 types of each specific bigram (collapsing the nearly merged ㅇ|| and ㅁ||), noted down their token frequencies, and divided the token frequencies by the overall tokens of the corpus.

Frequencies of word-initial German phonemes and bigrams were calculated using CLEARPOND for German (GermanPOND, Marian, Bartolotti, Chabal, & Shook, 2012), which is based on the SUBTLEX-DE Corpus, a corpus of movie and TV subtitles that is considered an excellent corpus for spoken German (Brysbaert et al., 2011). Austrian German differs from Middle/Northern German; however, the majority of German corpora include only small portions of Bavarian and/or Austrian varieties. In order to establish the applicability of the GermanPOND resource for Austrian data, the only available Austrian language corpus was compared to CLEARPOND to see whether Austrian and German lexical frequency rates are correlated and CLEARPOND can be used to analyze Austrian speech data. The ANNO Corpus of the Austrian National Library ("Austrian Newspapers Online," <http://anno.onb.ac.at>) is a collection of 20 million pages of Austrian newspapers and magazines published between 1527 to 2014. It is the only sizable corpus of Austrian German. There is a corpus of spoken Austrian German, the GRASS Corpus (Schuppler, Hagsmueller, Morales-Cordovilla, & Pressentheiner, 2014); however, it contains only spoken language and a limited number of speakers and tokens that can be analyzed with it. In the ANNO Corpus, the uninflected target words were searched between the

time period of 1950 and 2000 and the number of occurrences were noted down. As this corpus does not include a total token number but only gives the number of newspapers/magazines for a search period, token frequency was calculated per newspaper/magazine. For example, the word *Bank* (Engl. 'bank') occurred 652 times within the corpus, which was constituted of 3,124 newspapers and magazines for the respective time period. Frequency was calculated by dividing 652 by 3,124. This resulted in a lexeme frequency of 0.21 for *Bank*. Next, uninflected target words were searched in GermanPOND and their frequencies were extracted. The database underlying the German Clearpond calculators is the SUBTLEX-DE Corpus. The frequency values obtained from the ANNO and Clearpond corpora were z-scored, and then checked for correlations. They were correlated (Pearson's $r = 0.65$) and thus reliability of the GermanPOND resource for Austrian speech was assumed.

VOTs of EFL learners may be influenced by frequencies of items in the learned language. Thus, word-initial phoneme and bigram frequencies of EFL were calculated and compared to the frequency rates from the native languages. We used different EFL corpora from which we calculated the phoneme and bigram frequency rates for the EFL learners of Korean or German language background. For the Korean learners of English, the "ICNALE/ International Corpus Network of Asian Learners of English Corpus" (Ishikawa, 2013) was used. We calculated the frequency rate of word-initial phoneme and bigrams of the sub-corpus spanning only Korean learners of English by dividing the overall occurrences of the phoneme and bigrams by the number of tokens of the Korean corpus (= 246,879).

For the Austrian learners of English, data was extracted from two corpora, the "Louvain International Database of Spoken English Interlanguage" or LINDSEI (Gilquin, de Cock, & Granger, 2010) and the "Giessen-Long Beach Chaplin Corpus/GLBCC" (Jucker et al., 2006). We selected materials produced by speakers whose first language was German and determined initial phoneme and bigram frequencies by dividing the overall occurrences of the word-initial phonemes and bigrams by the corpus tokens (combined corpus size = 489,270).

CLEARPOND for English (Marian et al., 2012) was used to obtain English lexical frequency rates of the target words, initial plosive and initial bigram frequency rates. In addition, neighborhood density (i.e., number of phonological neighbors of the English target words differing by one phoneme) was calculated, as this variable plays an important role in lexical processing of first and foreign languages (Fricke, Baese-Berk, & Goldrick, 2016). Syllable frequencies were not calculated as the majority of word-initial syllables of the stimuli do not appear in Korean or German (e.g., 'bath,' 'dance'). This was due to the fact that many target words were monosyllabic (e.g., 'bills,' 'banks') and, thus, syllable frequency would be conflated with lexical frequency.

All phoneme and bigram frequency variables (L1, EFL, English) were first log transformed [$\text{LOG}(x+1)$] (to account for zero values in the data) and then rescaled to range between 0 and 1 in order to account for the different frequency distributions of phonemes and bigrams in the fortis and lenis category and per learner group. This allowed a direct comparison between Koreans and Austrians and between lenis and fortis consonants.

Statistical Analyses

First, a collinearity diagnostic was run on the independent variables (with the R packages “performance” and “car”) and correlation coefficients and variance inflation factors were computed (see Table 1).

Table 1

Correlation Matrix of Fixed Effects (Correlations Are Indicated in Bold)

	L1 phoneme frequency	L1 bigram frequency	EFL phoneme frequency	EFL bigram frequency	English phoneme frequency	English bigram frequency	English lexical frequency	English neighborhood density
L1 bigram frequency	-0.06							
EFL phoneme frequency	0.03	0.07						
EFL bigram frequency	0.14	0.32	0.29					
English phoneme frequency	0.22	-0.05	-0.27	-0.13				
English bigram frequency	-0.07	-0.02	0.001	-0.11	-0.46			
English lexical frequency	0.0	0.01	0.02	0.58	-0.12	-0.16		
English neighborhood density	0.03	0.03	0.02	0.08	-0.02	0.04	-0.23	
English neighborhood frequency	-0.02	0.003	0.002	-0.03	0.11	0.02	-0.21	-0.62

English phoneme frequency was shown to be correlated with English bigram frequency, and neighborhood density was correlated with neighborhood frequency. For each correlated pair, the first principal component (PC1) was computed via Principal Components Analysis in order to combine the two variables into one that can account for the majority of the variability of the two variables (Salem & Hussein, 2019). The first principal component (PC1) of “English phoneme frequency” and “English bigram frequency” was correlated negatively at -0.71 with each of the two variables and explained 75% of the data variability. The combination variable was termed “English phoneme/bigram frequency.” PC1 of “neighborhood density” and “neighborhood frequency” (termed “neighborhood density/frequency”) was correlated with each of the original variables at -0.7 and was able to account for 86% of the data variability.

A series of linear mixed models was then calculated (Bates, Maechler, Bolker, & Walker, 2014), with the dependent variable being the Mahalanobis distance scores of the learners and the fixed effects being (1) L1 phoneme frequency and (2) L1 bigram frequency. As control variables we entered (3) EFL phoneme frequency, and (4) EFL bigram frequency, (5) English phoneme/bigram frequency, (6) English lexical frequency rate, and (7) neighborhood density/frequency. As random effects (intercepts) we included ‘subject’ and ‘word.’ To keep type I error at the nominal level of 0.05, we included the maximal random slope structure (all fixed effects) per subject and per word (Barr, Levy, Scheepers, & Tily, 2013). Different models were computed with the Korean and the Austrian data.

As an overall test of the effect of the fixed effects, we compared the full model with a respective null model lacking the fixed effects (but being otherwise identical to the full model) using a likelihood ratio test (Dobson, 2002; Forstmeier & Schielzeth, 2011). We also tested the significance of individual fixed effects by comparing the full model with a respective reduced model lacking the effect to be tested. Due to low variance inflation factors, collinearity did not appear to be an issue (Field, 2005; Quinn & Keough, 2002). The models were implemented in R (R Studio Team, 2020) using the function *lmer* of the package *lme4* (Bates et al., 2014). The sample size for the models was 3,590 tokens, involving 86 types, and 43 speakers.

Figures were created with the R packages “interact” and “ggplot2.”

Results

American English speakers generally showed shorter VOTs before low vowels (see Table 2). The same pattern was true for Korean and Austrian learners of English, and these results are in agreement with previous literature on the influence of vowel height on VOT (Mortensen & Tøndering, 2013).

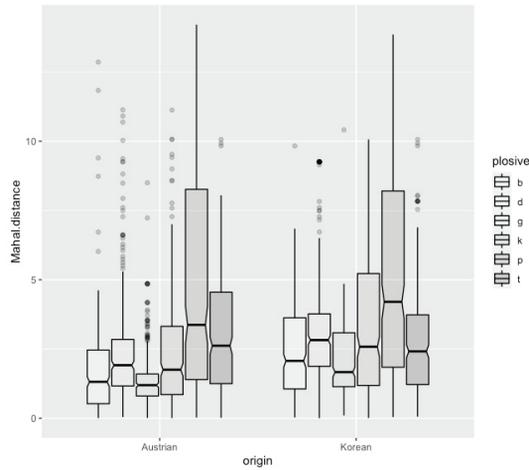
Table 2

Speech-rate-adjusted Lenis VOTs (in Milliseconds) of the American, Korean, and Austrian Speakers of English for Each Bigram (Means, Standard Deviations)

LENIS	B			D			G		
	ɑ, ʌ, a	e, ɛ, æ	i, ɪ	ɑ, ʌ, a	e, ɛ, æ	i, ɪ	ɑ, ʌ, a	e, ɛ, æ	i, ɪ
American English	53.1±39	75 ±23	116±32	84±11	109.8±38	177.8±55	118.2±37	293.3±218	160.3±43
Korean EFL	38±37	27±29	41±70	46±36	44±36	61±48	52±51	65±53	116±84
Austrian EFL	77±88	78±65	53±29	96±47	81±37	87±39	106±54	108±41	107±58
FORTIS	P			T			K		
	ɑ, ʌ, a	e, ɛ, æ	i, ɪ	ɑ, ʌ, a	e, ɛ, æ	i, ɪ	ɑ, ʌ, a	e, ɛ, æ	i, ɪ
American English	353±36	256±92	244±14	270.6±59	262.5±59	237.3±39.	262.9±29	243.2±41	263.7±53
Korean EFL	147±94	124±91	148±111	180±99	154±92	230±199	150±117	225±120	215±111
Austrian EFL	169±103	169±85	117±95	163±126	163±125	250±123	242±106	257±117	265±96

In total, 21.8% of Koreans' and 38.5% of Austrians' VOTs had a Mahalanobis distance of less than 1, which is close to the benchmark targets of the American English VOTs for their respective plosive types. Fortis plosives generally showed larger distances from the American English VOTs and the lenis plosives of the learners were closer to the American English phonetic spaces (see Figure 3). VOT distances of the lenis plosives were larger in Korean speakers, a fact that can be explained by the larger phonetic distance between Korean lenis and American English lenis VOTs. In addition, VOT distances of Koreans' /k/ also exceeded those of the Austrians. Both learner groups achieved the best VOT results for /g/. The largest Mahalanobis distances and variability in distances were measured for /p/ in both Koreans and Austrians.

Figure 3

Mahalanobis Distance per Plosive Type and First Language Background**Korean Results**

Results showed that Koreans' VOT distances were influenced by L1 bigram frequencies but not by L1 plosive frequencies (see Table 3 and Figure 4). Lower bigram frequencies facilitated smaller VOT distances to the American English model.

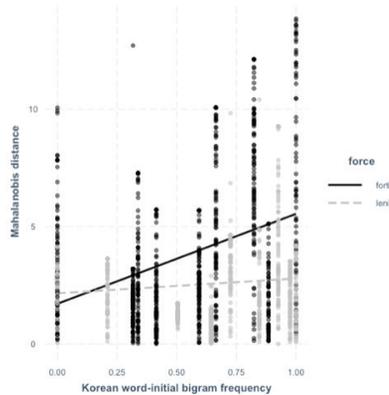
Table 3

Results of the Korean Linear Mixed Effects Models

Predictors	Estimate	SE	t	χ^2	p
(Intercept)	3.13	0.87	3.22		
L1 plosive frequency	-3.1	0.8	-3.9	1.21	0.27
L1 bigram frequency	0.46	0.1	5.7	8.48	0.004**
EFL plosive frequency	0.35	0.97	0.37	0.03	0.87
EFL bigram frequency	2.88	0.58	4.89	10.31	0.001**
English plosive/bigram frequency (PC1)	-2.93	0.43	-6.72	22.9	<0.001***
English lexical frequency	-0.15	0.14	-0.9	2.14	0.34
English neighborhood density/frequency (PC1)	0.13	0.24	0.56	0.29	0.58

Figure 4

Low Bigram Frequencies in Korean Facilitated Phonetic Learning in Word-initial Bigrams in Fortis and Lenis Plosives



In addition, EFL bigram frequencies and English plosive/bigram frequencies had an effect on VOT distances in the Korean learners (see Table 3), with the latter showing the opposite effect on VOT distances than L1 and EFL bigram frequencies: high-frequencies in the interaction variable of English plosives and bigrams led to smaller VOT distances in the Korean learners.

Austrian Results

VOT distances of the Austrian learners were affected by the frequency of the word-initial plosive in Austrian German (L1 plosive frequency), but not by L1 bigram frequencies (see Table 4 and Figure 5). High-frequency plosives showed more English VOTs than low-frequency ones.

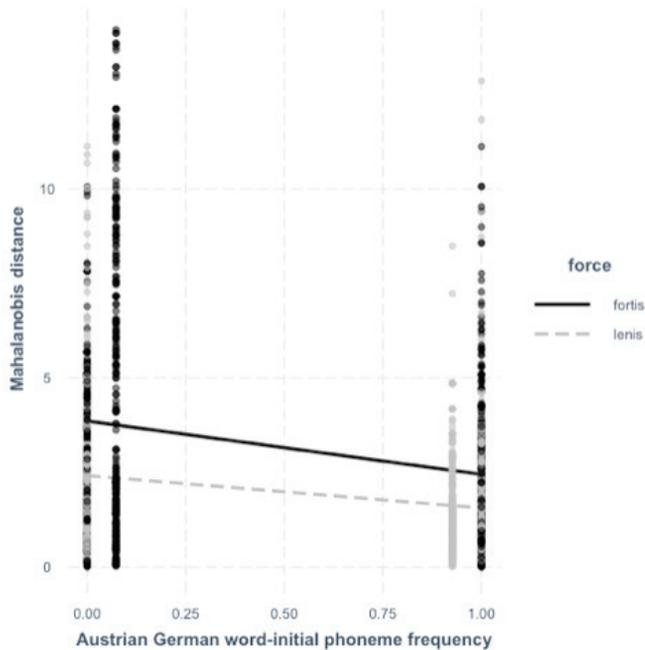
Table 4

Results of the Austrian Linear Mixed Effects Models

Predictors	Estimate	SE	t	χ^2	p
(Intercept)	2.95	0.27	10.9		
L1 plosive frequency	-1.89	0.24	-7.9	46.79	<0.001***
L1 bigram frequency	-12.1	6.5	-1.8	3.09	0.08
EFL plosive frequency	2.1	0.46	4.4	14.46	<0.001***
EFL bigram frequency	1.04	0.28	3.7	11.63	<0.001***
English plosive/bigram frequency (PC1)	-1.17	0.26	-4.57	21.54	<0.001***
English lexical frequency	-0.2	0.1	-2.02	8.82	0.003**
English neighborhood density/frequency (PC1)	0.13	0.15	0.88	6.16	0.013**

Figure 5

Austrian Learners Produced Better Approximations of American VOTs when German Phoneme Frequency of Fortis and Lenis Plosives Was High



EFL plosive and bigram frequencies also had an effect on Austrians' VOT distances, with low frequencies being indicative of shorter phonetic distances. High English plosive/bigram frequencies also had a measurable effect and minimized VOT distances. Words of high lexical frequency rate and words residing in sparser and lower frequency neighborhoods also showed improved VOT scores.

Discussion

The experiment conducted for the present study followed two investigative threads. First, we analyzed the role of phonotactic probability of initial phonemes (plosives) and phoneme combinations (bigrams: plosive plus vowel) on phonetic learning of voice-onset time in learners of English as a Foreign Language (EFL). Two competing hypotheses were tested: (1) high frequency rates of L1 segments slow down phonetic learning, and (2) high frequency segments have larger and more variable exemplar clouds, equipping

a speaker with more phonetic variability, and thus facilitating phonetic learning. We were specifically interested in analyzing the influence of the phonotactic probabilities that exist in the first language of EFL learners (Korean, German), as well as the influence of those probabilities formed through exposure to EFL of the two learner groups. Second, we tested whether sub-syllabic units play a role in phonetic learning and hypothesized that right-branching German syllable structure would influence phonetic learning of phonemes, while left-branching Korean syllable structure would influence the learning of bigrams. Thus, high-frequency German word-initial phonemes were expected to interfere with the learning of phonetic detail of equivalent structures in EFL in the Austrian group. In Koreans, high frequency rates of word-initial bigrams were proposed to be influential. The results show that frequency rates of word-initial segments were predictive in how far learners had progressed in their acquisition of English VOTs: high L1 frequencies affected phonetic learning in Austrian learners, while Korean learners were influenced by low L1 frequencies. Sub-syllabic segmentation was also shown to have an impact.

In general, the Austrian learners' English was influenced by a wider variety of factors analyzed in the present study. Neighborhood density and lexical frequency rate of target words were shown to have effects on VOT distances in Austrians but not in Koreans. The closer phonetic distance between English and German could play a role in this.

Concerning the first hypothesis, we found evidence that low-frequency items in the first language facilitate phonetic learning in English as a Foreign Language in Korean learners. In contrast, Austrians relied on high-frequencies to improve their English VOTs. These findings do not neatly fit into one of the proposed hypotheses. The Austrian results could be explained in the context of the exemplar-based hypothesis, where speakers have more numerous and diverse phonetic targets associated with high-frequency speech segments. When producing a novel sound in a foreign language, the Austrian learners may have a greater choice of phonetic patterns (or exemplars) for pronunciation. The Korean learners showed better VOT approximation to the American English model when frequencies of the respective segments in their L1 Korean were low. Here, the less automatized phonetic patterns associated with low-frequency bigrams may enable the phonetic learning process. The discrepancy between Austrians and Koreans could be related to the learning potentials that are different for each learner group. Austrians' VOTs were generally closer to the English model on the distance scale, whereas Koreans' VOT generally showed greater distances. When phonetic distances are small, the numerous phonetic competitors associated with the high-frequency segments could help hone in on the exact target. When phonetic distances are large, learners may have to ignore their L1 phonetic repertoire and acquire novel phonetic patterns in order to produce good approximations of a phonetic target. Low frequency rates

could facilitate that process, as they provide conditions where only a few and less deeply engrained phonetic targets exist, making it easier to adopt a new variant that is independent of the pre-existing phonetic variants.

The second hypothesis of sub-syllabic structure having an impact on phonetic learning in a foreign language was supported by our results. Due to left- and right-branched syllable structures differentiating the languages, we predicted Koreans to be mainly influenced by bigram frequencies, while Austrians to be mainly influenced by phoneme frequencies of their first languages. These expectations were borne out by the results, and Koreans' VOTs were shown to be affected by bigram frequencies of L1 Korean, whereas Austrians' VOTs were affected by L1 German plosive frequencies. The differences in cognitive linking of segments in language users' minds may be reflected in the differences in locus of frequency effects in EFL.

In sum, VOT distance reduction (i.e., more L1-user-like pronunciation of plosives) was most successful in cases where the first language probabilities of segments and segment combinations were low in Korean and high in Austrian German. Furthermore, in Koreans, distance reduction was largest when L1 Korean bigram frequency was involved, whereas in Austrians the reduction was largest when L1 German phoneme frequency was involved. This points to a role of sub-syllabic units in the cognitive processing of phonological features of a foreign language.

For better interpretation of the findings presented here, some limitations of the study should be considered. Carrier sentences differed in terms of subject phrase complexity and consequently higher rhythmic variability. In addition, a few cases of secondary stress on the initial syllable of a target word (such as in "punctuation" and "pizzerias") might have contributed to differences in VOT values. In general, the phonetics of VOT are heavily influenced by a variety of factors, including language experience (Stoehr, Benders, van Hell, & Fikkert, 2017), gender (Koenig, 2000), biological (hormonal) causes (Whiteside, Hanson, & Cowell, 2004), fluency of speech production (Beckman, Helgason, McMurray, & Ringen, 2011), and dialectal region of origin in Korea (Cho, 2005) and Austria (Moosmüller, 1987). In addition, the large inter-individual variation that is generally recorded in VOT measurements (e.g., Allen, Miller, & DeSteno, 2003) renders experimental designs complicated when trying to control for all of these factors. Future studies could compare L1 and L2 VOTs per person (paired data design) to document the exact VOT changes in a speaker switching from their first to their second language. A more detailed and separate investigation of the fortis and lenis categories may also yield interesting results that can qualify some of the findings presented here.

Conclusion

The results of the present study indicate that phonotactic probabilities in the first language exert influence over phonetic learning in a foreign language. Sub-syllabic structuring contributes to this effect by providing different segmental combinations where the frequency effects unfold.

In sum, our findings suggest an interaction between the statistical probabilities arisen in the first language, their cognitive entrenchment, and phonetic learnability in a foreign language, which is mediated by sub-syllabic segmentation of the first language.

Acknowledgements

The study was approved by the Internal Review Board of Seoul National University under IRB No. 1710/002-002. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (under grant NRF-2018S1A5A8028985 awarded to E. M. L.). We would like to thank Yaejin Jang and Jong-seung Sun without whose help this research would not have been possible. We are also grateful to Tomas Graf for assistance with the ESL corpora.

References

- Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75–86.
- Ahn, K. (2011). Conceptualization of American English native speaker norms: A case study of an English language classroom in South Korea. *Asia Pacific Education Review*, 12, 691–702.
- Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, 113(544). <https://doi.org/10.1121/1.1528172>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). `lme4`: Linear mixed-effects models using Eigen and S4. *R Package version 1.1–7*.
- Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, 39(1), 39–49.
- Berg, T., & Koops, C. (2010). The interplay of left- and right-branching effects: A phonotactic analysis of Korean syllable structure. *Lingua*, 120(1), 35–49.

- Berry, J., & Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *The Journal of the Acoustical Society of America*, 130, EL 365.
- Boersma, P., & Weenink, D. (2019). Praat, <http://www.praat.org> (Version 6.0.46).
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Boelte, J., & Boehl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412.
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–290.
- Bybee, J. (2007). *Frequency of use and the organisation of language*. Oxford University Press.
- Chang, C. B., & Kwon, S. (2020). The contributions of crosslinguistic influence and individual differences to nonnative speech perception. *Languages*, 5(4), 49.
- Cho, Y.-H. (2005). VOT and its effect on the syllable duration in Busan Korean. *Speech Sciences*, 12(3), 153–164.
- Chodroff, E., Godfrey, J., Khudanpur, S., & Wilson, C. (2015). Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*.
- Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, 54(5), 35–50.
- Dobson, A. J. (2002). *An introduction to generalized linear models*. Chapman & Hall/ CRC.
- Docherty, G. J., Watt, D. J. L., Llamas, C., Hall, D. J., & Nycz, J. (2011). Variation in voice onset time along the Scottish-English border. *Proceedings of the 17th International Congress of Phonetic Sciences*, 591–594.
- Espósito, A. (2002). On vowel height and consonantal voicing effects: Data from Italian. *Phonetica*, 59, 197–231.
- Everett, C. (2018). The similar rates of occurrence of consonants across the world's languages: A quantitative analysis of phonetically transcribed word lists. *Language Sciences*, 69, 125–135.
- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.
- Fischer-Jørgensen, E. (1980). Temporal relations in Danish tautosyllabic CV sequences with stop consonants. *Annu. Rep. Inst. Phonet. (University of Copenhagen)*, 14, 207–261.
- Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, 65, 47–55.
- Fricke, M., Baese-Berk, M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience*, 31(5), 639–645.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory & Language*, 42, 481–496.
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., & Dahlgren, N. (1993). *TIMIT: Acoustic-phonetic continuous speech corpus*. Linguistic Data Consortium.
- Gilquin, G., de Cock, S., & Granger, S. (2010). *Louvain International Database of Spoken English Interlanguage*. Presses Universitaires de Louvain.
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). *Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages*. Paper presented at the Proceedings of the 8th International Language Resources Evaluation (LREC'12).

- Grassegger, H. (1996). Koartikulatorische Einflüsse auf die Produktion von Anlautplosiven bei österreichischen (steirischen) Sprechern. In A. Braun (Ed.), *Untersuchungen zu Stimme und Sprache – Papers on speech and voice* (pp. 19–32). Stuttgart: Franz Steiner.
- Hunnicut, L., & Morris, P. (2016). Pre-voicing and aspiration in Southern American English. *University of Pennsylvania Working Papers in Linguistics*, 22, 215–224.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world* (pp. 91–118). Kobe, Japan: Kobe University.
- Jucker, A., S., M., & Smith, S. (2006). GLBCC (Giessen – Long Beach Chaplin Corpus). In *Oxford Text Archive*.
- Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45, 76–90.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832.
- Kim, J. (2015). Effects of phonotactic probabilities on syllable structure. *Working Papers in Linguistics*, 46(3), 1–16.
- Kim, J., & Davis, C. (2002). Using Korean to investigate phonological priming effects without the influence of orthography. *Language and Cognitive Processes*, 17, 569–591.
- Kim, J.-Y., & Lee, Y. (2011). A study of Korean syllable structure: Evidence from rhyming patterns in Korean contemporary rap-songs. *Sociolinguistics Journal of Korea*, 19, 1–22.
- Kim, J. Y. (2010). *L2 Korean phonology*. VDM.
- Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686–706.
- Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research*, 43, 1211–1228.
- Lee, H. B., Jin, N., Seong, C., Jung, I., & Lee, S. (1994). *An experimental phonetic study of speech rhythm in Standard Korean*. Paper presented at the International Conference on Spoken Language Processing (ICSLP), Yokohama, Japan.
- Levy, H., & Hanulíková, A. (2019). Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(9), 1–26.
- Lipani, L. (2019). Voice onset time variation in natural southern speech. *The Journal of the Acoustical Society of America*, 146(4). <https://doi.org/10.1121/1.5137433>
- Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16, 565–581.
- Luef, E. M. (2020). Development of voice onset time in an ongoing phonetic differentiation in Austrian German plosives: Reversing a near-merger. *Zeitschrift für Sprachwissenschaft*, 39(1), 79–101. <https://doi.org/10.1515/zfs-2019-2006>
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- 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), e43230.
- Martos, G., Muñoz, A., & González, J. (2013). On the generalization of the Mahalanobis distance. In J. Ruiz-Shulcloper & G. Sanniti di Baja (Eds.), *Progress in pattern recognition, image analysis, computer vision, and applications* (Vol. 8258). Springer.
- Moosmüller, S. (1987). *Soziophonetische Variation im gegenwärtigen Wiener Deutsch: Eine empirische Untersuchung*. Franz Steiner.

- Moosmüller, S., & Ringen, C. (2004). Voice and aspiration in Austrian German plosives. *Folia Linguistica*, 38, 43–62.
- Moosmüller, S., Schmid, C., & Brandstätter, J. (2015). Standard Austrian German. *Journal of the International Phonetic Association*, 45(3), 339–348.
- Morris, P. A. (2018). Rate effects on Southern American English VOT. *Proceedings of the Linguistic Society of America*, 3(60), 1–10.
- Mortensen, J., & Tøndering, J. (2013). *The effect of vowel height on voice onset time in stop consonants in CV sequences in spontaneous Danish*. Paper presented at the Proceedings of Fonetik 2013, Linköping, Sweden.
- Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 73–95). Springer.
- Phillips, B. (1984). Word frequency and the actuation of sound change. *Language*, 60(2), 320–342.
- Phillips, B. (2006). *Word frequency and lexical diffusion*. Palgrave Macmillan.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). John Benjamins.
- Port, R. (1983). Isochrony in speech. *Journal of the Acoustical Society of America*, 73, 66.
- Quinn, G. P., & Keough, M. J. (2002). *Experimental designs and data analysis for biologists*. Cambridge University Press.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>.
- Romani, C., Galuzzi, C., Guariglia, C., & Goslin, J. (2017). Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive Neuropsychology*, 34(7–8), 449–471.
- Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science*, 163, 292–299.
- Schoonmaker-Gates, E. (2015). On voice-onset time as a cue to foreign accent in Spanish: Native and nonnative perceptions. *Hispania*, 98(4), 779–791.
- Schuppler, B., Hagmueller, M., Morales-Cordovilla, J. A., & Pressentheiner, H. (2014). *GRASS: The Graz Corpus of Read and Spontaneous Speech*. Paper presented at the 9th edition of the Language Resources and Evaluation Conference, Reykjavik, Iceland.
- Schweitzer, K., Walsh, M., Calhoun, S., Schuetze, H., Moebius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech communication*, 66, 65–81.
- Shin, J. Y. (2008). Phoneme and syllable frequencies of Korean based on the analysis of spontaneous speech data (Seongin jayu ballwa jaryo bunseogeul batangeuro han hangueoui eumso mit eumjeol gwallyeon bindo). *Eoneocheonggakjangaeyeongu*, 13(2), 193–215.
- Shin, J. Y., Kiaer, J., & Cha, J. (2013). *The sounds of Korean*. Cambridge University Press.
- Siebs, T., de Boor, H., Moser, H., & Winkler, C. (1969). *Siebs deutsche Aussprache: Reine und gemäßigte Hochlautung mit Aussprachewörterbuch*. de Gruyter.
- Silva, D. J. (2004). Phonological mapping as dynamic: The evolving contrastive relationship between English and Korean. *Linguistic Research*, 21, 57–74.
- Silva, D. J. (2006). Variation in voice onset time for Korean stops. *Korean Linguistics*, 13, 1–16.
- Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *Phonetica Pragensia*, 2, 109–128. <https://doi.org/10.14712/24646830.2019.21>
- Sonderegger, M. (2015). Trajectories of voice onset time in spontaneous speech on reality TC. *Proceedings of the 18th International Congress of Phonetic Sciences*.

- Stoehr, A., Benders, T., van Hell, J. G., & Fikkert, P. (2017). Second language attainment and first language attrition: The case of VOT in immersed Dutch-German late bilinguals. *Second Language Research*, 33(4). <https://doi.org/10.1177/0267658317704261>
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321–1337.
- Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 49, 1175–1192.
- Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, 32, 827–853.
- Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics*, 14, 407–425.
- Todd, S., Pierrehumbert, J. B., & Hay, J. B. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, 185, 1–20.
- Vitevitch, M. S. (1997). The neighborhood characteristic of malapropisms. *Language and Speech*, 40, 211–228.
- Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1–16.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325–329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language*, 40, 374–408.
- Vitevitch, M. S., & Sommers, M. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production. *Memory & Cognition*, 31, 491–504.
- Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch range setting in voice acoustic analysis. *Behavior Research Methods*, 41, 318–324.
- Watt, D., & Yurkova, J. (2007). Voice onset time and the Scottish vowel length rule in Aberdeen English. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1521–1524.
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119(1), 597–607.
- Whiteside, S. P., Hanson, A., & Cowell, P. E. (2004). Hormones and temporal components of speech: Sex differences and effects of menstrual cyclicality on speech. *Neuroscience Letters*, 367(1), 44–47.
- Witzel, N., Witzel, J., & Choi, Y. (2013). The locus of the masked onset priming effect: Evidence from Korean. *The Mental Lexicon*, 8, 339–352.

Eva Maria Luef, Pia Resnik

Phonotaktische Wahrscheinlichkeiten und subsilbische Segmentation im Fremdsprachenerwerb

Zusammenfassung

Es ist bekannt, dass hohe phonotaktische Wahrscheinlichkeiten das Erlernen von Wörtern in der Erstsprache erleichtern. Die vorliegende Studie wurde konzipiert, um die Rolle phonotaktischer Wahrscheinlichkeiten beim Erlernen einer Fremdsprache zu untersuchen. Im Fokus standen österreichische und koreanische Englischlernende. Gegenstand der Untersuchung waren zwei Hypothesen, die mit phonotaktischen Frequenzeffekten in Zusammenhang stehen: (1) Hochfrequente Segmente haben tiefer verwurzelte phonetische Repräsentationen mit automatisierten Aussprachemustern, was das phonetische Lernen von homophonen Segmenten erschwert; (2) Hochfrequente Segmente sind mit einer höheren phonetischen Variabilität in der Erstsprache verbunden, was das phonetische Lernen in einer Fremdsprache erleichtern kann. Darüber hinaus wurde der Ort der Phonem-/Bigramm-Frequenzeffekte in Bezug auf die links- und rechtsverzweigte Silbenstruktur im Deutschen und Koreanischen analysiert. Dabei wurde festgestellt, dass die Nähe zur englischen Voice Onset Time mit den Phonem-/Bigramm-Frequenzen in der Erstsprache korreliert, allerdings variierten die Ergebnisse je nach Lernergruppe. Die subsilbische Segmentation der Erstsprache erwies sich ebenfalls als maßgebender Faktor. Die Studie stützt sich auf die Forschung zu Frequenzeffekten und kombiniert deren Grundannahme mit dem phonetischen Lernen in einer Fremdsprache. Die Ergebnisse zeigen einen engen Zusammenhang zwischen den statistischen Wahrscheinlichkeiten der Erstsprache und dem phonetischen Lernen in einer Fremdsprache.

Schlüsselwörter: Österreichisches Deutsch, Englisch als Fremdsprache (EaF), Frequenzverteilung, Koreanisch, subsilbische Segmentation

Appendix Table A1

Carrier sentences with sentence-initial plosives/bigrams

1. Touch screens are very useful nowadays.

2. Buffalos are large animals.

3. Gardening can be fun.

4. Tellers have to work long hours.

5. Gettysburg is a town in Pennsylvania.

6. Desk jobs can be boring.

7. Puff adders are very dangerous.

8. Tummy ache in little kids should not be underestimated.

9. Beavers live in lakes and rivers.

10. Gum ruins your teeth.

-
11. Cupboards in the kitchen need to be fixed.

 12. Deans of colleges have to work long hours.

 13. Pack horses have to be very strong.

 14. Cats are active at night.

 15. Gills of fish can look different ways.

 16. Customs is an agency responsible for collecting tariffs at the airport.

 17. Pucks are the balls of ice hockey.

 18. Death by car accident.

 19. Pictures of Tom can be found everywhere in this house.

 20. Text writing is a central feature of this class.

 21. Garry is his first name.

 22. Bathrooms are green nowadays.

 23. Tusks of elephants can be quite long.

 24. Battles of World War 2 included the one at Normandy.

 25. Duffel bags are convenient for travelling.

 26. Passion for sports runs in my family.

 27. Kitties are little cats.

 28. Dish washers are too expensive for me.

 29. Kerosene is fuel for jet engines and lamps.

 30. Geese can swim.

 31. Bees make honey.

 32. Dust gathers easily in the corners of apartments.

 33. Bats live in hollow trees.

 34. Tim is my brother.

 35. Peanuts can be bad for your health.

 36. Dance balls are old-fashioned.

 37. Custard recipes are typically milk-based.

 38. Buck is his nickname.

 39. Deals in the business world are hard to make.

 40. Gifts are given for Christmas.

 41. Tea ceremonies are known from Japan.

 42. Guesswork is the process of making a guess when you do not know all the facts.

 43. Pumpkins are my favorite vegetable.

 44. Dusk is the time before the sun rises.

 45. Buddy systems for language learning are a great invention.

 46. Gut microbes are important for your health.

 47. Pieces of the cake are in his hair.

 48. Cutlery can be bought at the supermarket.

 49. Tins have to be recycled.

-
50. Pillows can be expensive in this store.
-
51. Geckoes are little reptiles.
-
52. Deserts are defined as dry lands.
-
53. 'Pancake-House' is open today.
-
54. Kings of England.
-
55. Buns for burgers can be very soft.
-
56. Guns are used for killing people.
-
57. Punch contains a lot of sugar.
-
58. Teak wood comes from the rainforest.
-
59. Bundles of joy.
-
60. Guests are not welcome at my house.
-
61. Pills are generally prescribed by your doctor.
-
62. Differences in opinion should not be expressed.
-
63. Punctuation marks need to be inserted.
-
64. Telegrams are not used anymore nowadays.
-
65. Chemicals in your clothes are bad for your skin.
-
66. Bishops work for the church.
-
67. Tussles should be avoided!
-
68. Cans have to be recycled.
-
69. Bills just keep piling up.
-
70. Gutters can be found on the street.
-
71. Ducks live in lakes and ponds.
-
72. Tennis players need to have strong muscles in their arms.
-
73. Cups you can find in the upper left shelf.
-
74. Pepper can be spicy.
-
75. Duds are expensive to buy.
-
76. "Killerbird" is the name of a movie.
-
77. Ticks carry lots of diseases.
-
78. Pets are not allowed in the apartments.
-
79. Kids have to go to school.
-
80. Dill is an herb used for Italian cooking.
-
81. Beach houses were affected by the hurricane.
-
82. Tests will not be written this semester.
-
83. Gears in the car are for shifting.
-
84. Banks reliably store your money.
-
85. Decks of cards.
-
86. Kiss for you, kiss for me.
-

Appendix Table A2

**Corpora (T = TIMIT, TAL = This American Life), carrier sentences,
and speaker information of American speakers**

Nbr.	Bigram	Corpus	Sentence	Nbr. of speakers (m, f)	Mean speaker age
1		T	Barb's gold bracelet was a graduation present.	1, 1	31
2		T	Bob found more clams at the ocean's edge.	1, 0	28
3	b + a, ʌ, ɒ	T	Bob papered over the living room murals.	4, 1	34.6
4		T	Barb burned paper and leaves in a big bonfire.	3, 3	29.8
5		T	Butterscotch fudge goes well with vanilla ice cream.	1, 0	30
6		T	Bagpipes and bongos are musical instruments.	5, 2	31.7
7	b + e, ɛ, æ	T	Beverages are made from seeds the world over.	1, 0	27
8		T	Basketball can be an entertaining sport.	3, 3	29.8
9		T	Beer, generally fermented from barley, is an old alcoholic beverage.	1, 0	27
10	b + i, ɪ	T	Biblical scholars argue history.	3, 4	25.3
11		TAL	Beers are \$2.50.	1, 0	40
12	d + a, ʌ, ɒ	T	Ducks have webbed feet and colorful feathers.	7, 0	35.1
13		T	Death reminds man of his sins.	0, 1	24
14		T	Dances alternated with sung or spoken verses.	0, 1	35
15		TAL	Dad, are you doing OK?	0, 1	41
16	d + e, ɛ, æ	TAL	Dad, I'm so sorry I always used to say you were stinky.	0, 1	41
17		TAL	Dad?	1, 0	50
18		TAL	Dan was born in South Bend in 1946, same year as the club.	1, 0	40
19		TAL	Dan told me he thinks that it wasn't what Obama said.	0, 1	39

20		T	Differences were related to social, economic, and educational backgrounds.	0, 1	25
21		TAL	Deanna got a postcard of him that year when her family went to Universal.	0, 1	33
22	d + i, ɪ	TAL	Deanna called her Aunt Rose from the basement, distressed.	0, 1	33
23		TAL	Dishwasher Pete, a real live dishwasher.	1, 0	38
24		TAL	Dish out of water.	1, 0	38
25		TAL	Dishwashers are invisible to most restaurant customers.	1, 0	30
26	g + a, ʌ, ɒ	T	Gus saw pine trees and redwoods on his walk through Sequola National forest.	4, 3	35.7
27		TAL	Gamblers in Dixon's lab will inevitably say that the near misses are closer to a win than a loss.	0, 1	43
28		TAL	Gambling's wrong.	1, 0	55
29	g + e, ɛ, æ	TAL	Gary did not want to become a football player.	1, 0	60
30		TAL	Gary is a comedian today.	1, 0	60
31		TAL	Gary, they will kill you.	1, 0	49
32		TAL	Ghetto hoochie mama.	1, 1	27
33		TAL	Geese were on the other side of this area when I was talking.	1, 0	42
34	g + i, ɪ	TAL	Geese are nasty.	0, 1	64
35		TAL	Geeks move in.	1, 0	38
36		TAL	Geese are laying.	1, 0	56
37		T	Publicity and notoriety go hand in hand.	3, 3	25.5
38	p + a, ʌ, ɒ	T	Palm oil protects the surfaces of steel sheets before they are plated with tin.	1, 0	44
39		T	Pa don't care about the kid.	0, 1	26
40	p + e, ɛ, æ	T	Penguins live near the icy Antarctic.	5, 2	30.5
41		T	Pam gives driving lessons on Thursdays.	1, 2	40
42	p + i, ɪ	T	Pizzerias are convenient for a quick lunch.	5, 2	32.8
43		T	People never live forever.	1, 0	25
44	t + a, ʌ, ɒ	T	Tugboats are capable of hauling huge loads.	5, 2	28.1
45		T	Todd placed top priority on getting his bike fixed.	6, 1	29

46	t + e, ε, æ	T	Technical writers can abbreviate in bibliographies.	6, 1	29.7
47		T	Tetanus could be avoided by pouring warm turpentine over a wound.	0, 1	27
48	t + i, ɪ	T	Tim takes Sheila to see movies twice a week.	2, 5	32
49		T	Teaching guides are included with each record.	0, 1	26
50		T	Carl lives in a lively home.	7, 0	30.8
51	k + a, ʌ, ɒ	T	Cottage cheese with chives is delicious.	4, 2	26.5
52		T	Coffee is grown on steep, jungle-like slopes in temperate zones.	4, 3	29.2
53		T	Calcium makes bones and teeth strong.	4, 3	37.1
54	k + e, ε, æ	T	Castor oil, made from castor beans, has gone out of style as a medicine.	1, 0	45
55		T	Cattle which died from them winter storms were referred to as the winter kill.	0, 1	28
56	k + i, ɪ	T	Kindergarten children decorate their classrooms for all holidays.	6, 1	34.4