SOCIALLY-CONDITIONED LINKS
BETWEEN WORDS AND PHONETIC REALIZATIONS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITYOF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR IN PHILOSOPHY

IN

LINGUISTICS

May 2018

By

Jonny Kim

Dissertation Committee:

Katie Drager, *Chairperson*
Amy Schafer
Theres Grüter
William O’Grady
Abby Walker
Sang Yee Cheon

Keywords: lexical recognition, sociophonetics, word characteristics, talker information

Dedicated to all creative ones

who are leading and participating in

linguistic change

ACKNOWLEDGEMENTS

When I was a lad, ‘a doctor’ meant to me a person who knows everything. The more I learn, the more certain it becomes that there is always more to learn. One thing I certainly know, though, is that I am blessed with all teachers I’ve had in my life. There are so many who’ve showed me the way to where I am that I can’t list them all here. I would like to express the deepest gratitude to my dear advisor, Katie Drager. You’ve shaped my academic identity, my thoughts, and my beliefs. And, I owe many thanks to my committee members: Amy Schafer, Theres Gruter, William O’Grady, Sang Yee Cheon, Abby Walker. I can’t thank enough for the encouragement and inspiration they gave. I can’t also imagine what my grad life would have been like without professors in the LLL College who taught and worked with me. My dissertation work was largely motivated by Abby Walker and Jen Hay. I deeply appreciate their insight. My work also benefitted from insights of Paul Foulkes, Holger Mitterer, Bradley Rentz, and anonymous reviewers of LabPhon and TopiCS.

My research and life in Hawai‘i would have not been sustainable without sincere help of the LING staff members (Jen, Justin, Cheri, Nora), and financial support of the Bilinski Educational Foundation and the Hawai‘i Department of Human Services. My research is also indebted to my “socially-appropriate talkers”: Inso Yi, Munja Kim, Wooyoung Cho, Grace Hwamok Cho, Yuri Lee, Eunyoung Song, Hankyul Kim, and Dongjin Lee. I am appreciative to their voices even when they mismatch word age. There were more volunteers who willingly provided their voice samples and many thanks also go to many participants who spent time and efforts in my experiments. I also would like to thank Dr. Jaeho Yoo and Akiko Kahue-Burrows at CAPE, Giroung Lim, Bonnie Fox, and the LBC program for providing access to participant

pools in Hawaiʻi. I also would like to express thanks to Kirt and Aitor at CLT who provided lab-technical support.

Grad students in the LLL College filled my research life with happiness. I thank all my friends in the LING department: Yuri (you deserve being mentioned twice), Jinsun, Chaeeun, Onsoon, Sejung, Sunju, Kumjeong, Gyu-Ho, Tyler, Amber, Melody, James, Joelle, Nozomi, Claire, Thomas, Jessica, Crystal, John, David, Katie, Elaine, Peter, Kirsten, Ryan, Catherine, Tobias, Victoria, Aaron, Kevin, Sharon, Jake, Kavon. Also big thanks to my friends in SLS and EALL: Hyunah, Hyunwoo (& JiHyun), Kitaek, Hye Young (Jung), Eunseok, Sangki, Rue, Wenyi, Aya, Nancy, Soyoung, Eunho, Hyeyoung (Choi), Sooran, Sumire, Rachel, Reed. I also appreciate the Korean professors at UHM who have supported my research and life during the course: Ho-min Sohn, Sumi Chang, Sun Hee Park.

Dr. Sunhoi Kim and professors at Chung-Ang University have always supported me and reminded me of the happiest time when I first began dreaming of a linguist. I am lucky that I have friends who share those happiest memories: Hankyul (you also deserve twice), Jonggyeong, Gwan-Hyeok, Gina, Yeori, Jungeun. I was also lucky to have people who supported me with data collection at Chung-Ang: Sejin, Hana, SungHyun, Ilsoo, Yoomin, Dongyeong, Sunmin, Dayoung, Gyunam. My old friends, Sunhan, Ik, Sungmin, also supported me in various ways during data collection in Korea.

People in Hawaiʻi were awesome! I made full use of the energy that Sam at Glazers Coffee gave everyday. I will remember the joyful times with my baseball buddies: Seunghyun, Jeongmin, Sean, Andrew, Greg, Cam. My family and I spent many unforgettable moments with my grad-scientist buddies, Ilsoo, Jinhee, Hyewon from the first days we settled in Hawaiʻi. Mahalo to be a linguist! All memories in Hawaiʻi will be cherished in my family's entire life.

I am certain that my father is proud of his two sons in heaven. My brother Jungho's and my families will keep going on this journey depending on each other. Dad, we didn't have much time to talk while you were here, but now I have so much I want to tell you (how much I love you). Mom, you may not have much time to see while I am with you, but I have so much I want to show you (how much I love you). Last, to my love Hye Jin Yoon, Dana, Lena (& Sona), I am thankful every moment for your devotion. I love you more than words. I am with you all my life.

ABSTRACT

Previous research on speech perception has accumulated evidence for a claim that phonetic detail in previously encountered utterances is stored in lexical memories, and speech processing benefits from listeners' probabilistic knowledge about distributions of word-specific phonetic patterns over social categories of the speakers. Building on that claim, this dissertation explores the degree to which word recognition is informed by the experience-based links between phonetic and lexical information through experiments using Korean words and phonetic realizations indexed to different age groups.

In Chapter II, two lexical decision experiments replicate Walker and Hay's (2011) finding that lexical access is improved when the word is produced by a talker from the age group who produce the word most frequently. Further, Experiment 1 demonstrates that the effect of age-congruent realizations is enhanced when the word is stereotypically associated with age groups, beyond the effect of distributional associations between words and age groups. In Experiment 2, the effect arose even when listeners held no expectation about the talker prior to the word onset, suggesting that lexical access is rapidly boosted by socio-indexical phonetic cues that are congruent with socio-indexical information of the word.

In Chapter III, another lexical decision experiment (Experiment 3) provides evidence that exposure to a single socially-indexed phonetic variant – as opposed to target words produced by varying talkers – is sufficient to prime words that are associated with similar social information. Words were recognized faster when the word was preceded by a prime word that contained a phonetic variant associated with the age group that the word is associated with. However, the effect did not occur when the prime word produced by the same talker contained a phonetic variant that is not associated with age, suggesting that the priming process may not require explicit awareness of the current talker's identity or activation of abstract representations of age-related information.

In Chapter IV, an eye-tracking lexical identification experiment (Experiment 4) tested whether processing during phonetic ambiguity at the word onset was affected by age-indexed information of the word and the talker, and this was tested while two orthographic word candidates and the talker's voice were provided prior to the auditory stimuli. It was hypothesized that a candidate word associated with the same age with the talker would be fixated more frequently before the target word was disambiguated, which would provide evidence that lexical access is consulted in real time by sociophonetic detail encoded in lexical representations. However, an unexpected pattern was found; more frequent fixations and faster identification responses were observed for words produced by age-incongruent talkers in a time region following the retrieval of phonetic disambiguation cue. Given the presence of pre-activated age-related information from the words and the talker, the results demonstrate listeners' preparatory attention to, and strategic use of, prior information about the socially-conditioned associations in an effort to overcome the predicted socially-incongruent phonetic realizations of the words.

The overall results are consistent with predictions of experience-based cognitive mechanisms of language processing; memories and recognition processes of phonetic and lexical information are jointly shaped by listeners' socially-conditioned experience with phonetic variation and lexical use, and listeners can selectively adapt to sociophonetic variability and contextual information in accordance with communicative purposes. Further issues that the results raise will be discussed, including empirical questions that need to be addressed to better understand the role of socio-indexical phonetic detail in speech perception.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

Since the emergence of sociophonetics in the 1970s, the area delved into the study of phonetic variation with a focus on the relationship between phonetic forms and social factors. Experimental work in sociophonetics has accumulated evidence that phonetic details are remembered and indexed to social information, and that listeners use these links during phonetic processing (Johnson, 2005; Hay & Drager, 2007). On the other hand, indexical properties of lexical items and their interplay with sociophonetic realizations are relatively understudied. This dissertation explores the underlying mechanism of spoken word recognition based on results of a series of experiments.

Using various psycholinguistic paradigms, the experiments examine recognition facility (i.e., accuracy, response time, eye fixation rate) of Korean words associated with different ages when they are produced by talkers of varying ages or talkers producing different age-indexed phonetic realizations. The main purpose is to test the degree to which listeners are influenced by a combination of social information indexed to the lexeme and the acoustic cues in the signal. The experiments provide evidence that lexical processing is rapidly and automatically informed by social information encoded in the phonetic detail. The results are discussed in light of existing models of spoken word recognition.

In this introductory chapter, I will present a review of relevant research coming from spoken word recognition and sociophonetics, as well as their related disciplines, holding on to an experience-based perspective that human cognition is wired in a way that listeners can cope with the messy variation of spoken language through a probabilistic inference process. In this view, listeners resolve transient uncertainty of the utterance by incrementally collecting and integrating evidence including socio-indexical information of phonetic forms and lexical items.

The review begins with an overview of key assumptions made in the earlier framework of spoken word recognition during the latter half of the twentieth century, focusing on the time course of lexical access and exchange of information between the lexical and acoustic/phonetic levels (Section 1.1). The next section discusses diverging assumptions about the nature of lexical representations, pointing out the neglect of listeners' socially-conditioned experiences in traditional accounts of speech perception, namely *abstractionism*[1] (Section 1.2). Next, I will discuss previous findings on listeners' ability to use socio-indexical information during speech recognition in Section 1.3, highlighting the need for a theoretical frame integrating sociophonetic influences. In Section 1.4, I will examine how listeners' socially-conditioned experience with phonetic variation can influence lexical-level memory and processing. In Section 1.5, experience-based models of speech processing (exemplar-based and Bayesian-based models) are discussed, focusing on how social and phonetic aspects of the speech signal are integrated in the cognitive system. Finally, I will give an overview of the dissertation by summarizing the research questions addressed in each chapter (Section 1.6).

1.1. Time course of spoken word recognition

While comprehending spoken language, we integrate a vast amount of acoustic-featural information on the fly. Given the temporal constraints of the speech signal, the rapidness and automaticity of the perceptual integration raise a number of questions for psycholinguists. How

---

[1] I restrict using this term to refer to the rule-based theory that presupposes abstract and invariant representations as well as fixed access routes (e.g., generative models and normalization models).

do we map the auditory input with a word in the lexicon? Do we process the low-level acoustic-phonetic information to recognize the phonemes first, and then integrate the phonological analysis with lexical knowledge? What are the roles of probabilistic factors of the word? Essentially, how rapidly does our perceptual system consider prior knowledge about the word (e.g., semantic meanings, intervention of auditorily similar words, probabilistic information of the word)? This section reviews general principles of spoken word recognition with these questions in mind, relating to the temporal aspect of auditory perception on the one hand and the directionality of information flow between the phonetic and lexical levels on the other hand.

Current models of spoken word recognition are grounded on the following theoretical consensus made during the initial stage of the field (Weber & Scharenborg, 2012; Magnuson, Mirman, & Myers, 2013; Mattys, 2013). First, phonemic inputs activate multiple word candidates stored in the mental lexicon in parallel, and the activated words compete each other for recognition. Second, the degree of activation is determined based on goodness of acoustic fit between the incoming signal and the lexical representation. Controversies in the field boil down to two major problems; (1) the timing of an influence from top-down contextual information on lexical access (i.e., whether the initial mapping process between the signal and the word form is susceptible to top-down contextual information), and (2) the nature of lexical representations (i.e., whether a word is represented as a single abstract representation or a more phonetically-detailed format, such as distribution of episodic traces).

This section focuses on diverging viewpoints for the first issue. The second issue will be discussed in the following sections with respect to factors contributing to socially-conditioned speech recognition. To anticipate the literature review in the later sections, theoretical perspectives for these issues are enriched by integrating the roles of sociophonetic factors in phonetic processing of spoken words. In the following chapters, further discussion on the sociophonetic framework in regard to these issues will be provided, in the context of the results from the experiments in this dissertation and suggestions for future work.

1.1.1. Simultaneous integration of low-level phonetic cues

An important constraint in the temporal aspect of spoken word recognition is that the auditory signal is transient and time-bound, so we only retrieve linguistic information from the acoustic content in a sequential fashion. Marslen-Wilson and colleagues (e.g., Marslen-Wilson & Tyler, 1980) suggested that perceptual interpretations of the unfolding speech take place in real time as left-to-right analysis for a string of phonemes, and this became a key assumption for earlier models of spoken word recognition. This assumption entails a strong constraining role of word onsets, such that multiple words consistent with the observed onset begin to be activated and compete for recognition immediately upon hearing the initial portion of an utterance. This idea is supported by empirical findings using various experimental methods.

For example, participants in Grosjean's (1980) gating task[2] provided a variety of possible completions for the target word based on the earlier potion of the input, indicating auditory-based parallel competition of multiple candidate words. However, individuals' responses tended to converge on the most likely completion of the target word without hearing the word in its

---

[2] The gating paradigm is a time-course method developed for speech perception research in its infancy. In Grosjean's (1980) original design, participants repeatedly listened to a fragment of speech, each time increasing the length of the fragment (e.g., the initial consonant of the target word for the first "gate", and then the initial CV syllable for the second gate). And then, they guessed what the complete speech material (i.e., the target word in its entirety) would be and gave their subjective confidence rating.

entirety, suggesting that phonetic cues in the word onset provide key evidence in predicting the linguistic content that follows.

While such results may be induced merely by phonemic or featural match, the evidence that a cohort of multiple words (i.e., words beginning with the identical phonemes) are activated at the lexical level comes from cross-modal priming experiments, which utilize activations via semantic links between lexical items. Marslen-Wilson and Zwitserlood (1989) and Zwitserlood (1989) demonstrate that hearing a priming word (e.g., *captain*) activates its cohort (e.g., *capital*) and then the activation spreads to facilitate recognition of written words representing semantic relatives of either the target (e.g., ship) or the cohort word (e.g., *money*). However, the priming effect was not found for semantic relatives of a word that rhymes with the priming word (e.g., *Captain* may activate *mountain* but recognition of *forest* was not facilitated), highlighting the dominant role of auditory similarity in word onsets during lexical competition, compared to that of the later part of the word.

Further, experimental paradigms measuring online processing provide more direct evidence that lexical competition is led by integration of onset cues in real time. Numerous studies using the visual world paradigm – in which listeners' eye movements on visual stimuli are tracked while perceiving speech materials (see Section 4.1.1 for more detail about this paradigm) – have shown that an object or a string of letters representing a cohort word is fixated (therefore considered as a potential target) as frequently as the target object during the initial time period under phonetic ambiguity (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan et al., 2001b; Dahan & Tanenhaus, 2004).

As the first psycholinguistic model of spoken word recognition, the Cohort model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980) provides an efficient mechanism for the rapid bottom-up integration of phonetic cues into lexical activation. According to this model, lexical recognition takes place in three stages: *access*, *selection*, and *integration*. In the access stage, acoustic patterns of the signal are mapped onto word forms in the lexicon, and all words consistent with the phonetic cues are activated immediately, as early as 150-200ms (approximately the duration of the first two segment in a normal speech rate). During selection, activated word forms are assessed by acoustic similarity to best match the incremental phonetic cues. As more acoustic input is retrieved, mismatching words are continuously removed from the cohort set, ideally until the uniqueness point, the point at which only one word becomes uniquely identifiable. Lastly, during integration, the words remaining in the cohort set are assessed by the context (i.e., semantic/syntactic representations of the word).

The original Cohort model was a partially-interactive model that allowed top-down feedback from the context during selection, but the model was revised to a fully bottom-up model (Marslen-Wilson, 1987; 1989), in which the access and selection occur only as a form-based process, and words mismatching the top-down information are removed from the cohort in the integration stage. Thus, in line with other models with a bottom-up priority – e.g., Race (Cutler & Norris, 1979), FLMP (Massaro, 1987; 1996), Merge (Norris, McQueen, Cutler, 2000), Shortlist (Norris, 1994), Shortlist B (Norris & McQueen, 2008) – Cohort posits that prelexical phonological analysis (i.e., the access stage) is not corrupted by lexical information (or even higher-level information), but interactions with top-down information occur as a consequence of post-perceptual assessment.

1.1.2. Simultaneous integration of high-level lexical information

In the previous section (1.1.1), we reviewed evidence for activation and competition of multiple lexical items initiated by simultaneous integration of bottom-up acoustic cues given at the early portion of the word. On top of that, spoken word recognition utilizes information about the word itself. Early evidence for lexical intervention includes (1) the word superiority effect (Reicher, 1969; Rubin, Turvey, & Van Gelder, 1976), where phonemes are identified faster when embedded in a real word than in a non-word, (2) the phoneme restoration effect (Warren, 1970) where a phoneme segment auditorily masked by extraneous noise is not noticed as missing but listeners instead restore it as a phoneme that is consistent with the lexical/sentential context, (3) the Ganong effect (Ganong, 1980) where a speech segment ambiguous between two phonemes tend to be perceived as one that constitutes a real word given its phonological context. These effects allude a right-to-left influence of context, and recent studies using real-time measurements provide evidence for immediate integration of lexical knowledge, as in the Ganong effect (Kingston et al., 2016, using the visual world paradigm) and an effect of semantic properties (Zhuang et al., 2011, using fMRI neuroimaging).

In addition, as for an effect of representations higher than the lexical level, Marslen-Wilson's (1975) sentence shadowing task showed that interpretations of the signal are immediately integrated with higher-order structural analysis beyond the lexical level, which is also supported by visual world paradigm (Magnuson, Tanenhaus, and Aslin, 2008).

Crucially, lexical recognition is influenced by prior probability of words, such as word frequency. Numerous studies have demonstrated that high-frequency words are perceived with higher recognition facility (Luce, 1986; Marslen-Wilson, 1987, 1990; Connine, Titone, & Wang, 1993; Luce & Pisoni,1998; Dahan, Magnuson, & Tanenhaus, 2001a; Dufour, Brunelliere, &

Frauenfelder, 2013). Fox (1984) also showed a word frequency effect on phonemic distinction, in which identification of a word-onset segment that is auditorily ambiguous between two words with varying frequency (e.g., between a high-frequency word *best* versus a low-frequency word *pest*) is biased toward a phoneme consistent with the high-frequency word.

The aforementioned cross-modal priming effects are also modulated as a function of frequencies of the target and competitors. As for the temporal aspect of the frequency effect, the frequency effect on the priming tasks is reported to occur only temporarily when the visual target is provided before the auditory uniqueness point, and the effect disappears when the visual target is presented in the later phase (Tyler, 1984; Zwitserlood, 1985). In a lexical decision experiment (see Section 2.1.1 for the details about this method), response times for monosyllabic words are influenced by word frequency, but not for disyllabic words because longer words provide sufficient time for the frequency effect to fade out (Blosfeld & Bradley, 1981). Likewise, in McQueen's (1991) phoneme decision test, fast responses were influenced by lexical frequency to a greater degree than slow responses. The early impact of frequency on lexical access is also supported by real-time measurements using eye-tracking (Dahan et al., 2001a) and event-related potentials (ERPs) (Dufour et al., 2013)[3].

In accordance with these findings, Marslen-Wilson (1987, 1989) proposed that the Cohort mechanism code frequency as a predictor for the resting activation rate of a lexical item (i.e., the baseline activation level of a word). However, in spite of its efficacy to explain the temporal aspect of the immediate acoustic-lexical mapping, the Cohort theory encountered

---

[3] See, however, Balota and Chumbly (1984) and Connine et al., (1993) for a diverging viewpoint about the temporal locus of the word frequency effect. In both studies, the authors claim that the word frequency effect (traditionally demonstrated by non-real-time methods) occurs at a post-perceptual phase due to task-specific decision biases.

criticism mainly for its inability to account for the fact that listeners recognize words that indeed mismatch the onset (especially in noise) and lexical activation derived from similarity in the later part of the word.

A mathematical model, Neighborhood Activation Model (NAM) (Luce, 1986; Luce & Pisoni,1998) uses a different metric of acoustic similarity between the input and lexical competitors than a cohort competitor set. In this model, competition occurs among phonological neighborhoods of the spoken word; that is, any words that differ from the input by deletion, addition, or substitution of one phoneme regardless of its position within the word. The activation rate of a candidate is estimated by frequency-weighted neighborhood probability, rather than the onset match.

The time course of competition among words that match any part of the input is explicitly predicted by a computationally implemented model, TRACE (McClelland & Elman, 1986). In contrast to the sequential feed-forward models adopting bottom-up inhibition of lexical candidates, TRACE is a connectionist model that assumes that simultaneous integration of lexical feedback is necessary to predict the right-context effect. Such a mechanism is referred to as the interactive activation framework and is realized by a network composed of excitatory connections between levels and inhibitory connections among lateral units. Via the excitatory link, the input of multidimensional auditory features temporarily activates their associated phonemes, and the phoneme-level activation is again fed forward to the word level. At the same time, lexical competition is realized via parallel inhibitions; as activation of a node increases, other competitor nodes are inhibited by the dominant node.

Crucially, the links between phonemes and words inform one another and continuously mapped based on the overall similarity of the acoustic memory traces aligned at time slices.

Since activation rates depend on repetitive measure of the overall similarity (rather than complete match of the onset), the model is more lenient (than Cohort) for uncertainty of the input and can predict recovery from confusability. In such a mechanism that handles relative activations based (partially) on fine-grained acoustic fit, recognition emerges (rather than takes place at a specific time point) when the lexical node with the highest inhibitory weight dominates the lexical level. An advantage arising from such implementation is that TRACE predicts different time courses of activation for cohort and rhyme competitors; while cohort words are activated to a greater degree than rhyme words in the initial stage, the rhyme activation becomes dominant at the end due to increasing similarity. However, the peak activation rate for rhyme competitors is never higher than the cohort competitors, predicting stronger inhibition of the target and cohort than rhyme activation.

These patterns are in accordance with what Allopenna, et al. (1998) observed from human perception, which provided the first real-time evidence for activation of competitors whose onset does not correspond to the signal. In their eye-tracking experiment, listeners were instructed to move an object in the visual scene by mouse drag, while the visual scene presented pictured objects representing (1) the auditory target (e.g., a *beaker*), (2) a cohort competitor (e.g., a *beetle*), (3) a rhyme competitor (e.g., a *speaker*), and (4) an unrelated item (e.g., carriage). As briefly mentioned above, the fixation rate of cohort competitors was equivalent to the target (and higher than the unrelated baseline items) in the initial stage. Rhyme competitors also began to draw more frequent eye fixations than the unrelated items slightly later but the peak fixation rate did not exceed the target and the cohort. This finding opened a novel prospect for the field in support of continuous mapping models, such as TRACE. That is, the time-locked roles of cohort and rhyme competitors suggest that lexical access emerges from relative activation of lexical competitors in relation to temporal memory traces of the input, reconciling the influences of overall similarity (as predicted by NAM) and congruence in the word onsets (as predicted by Cohort).

1.2. The nature of lexical representations

While the previous section discussed how speech sounds are mapped onto words with regard to the temporal constraint, another challenge as a listener is to overcome variability of the signal. Human perception often includes mapping the sensory input for outer stimuli that are gradient by nature into a set of discrete categories stored in the mind. With respect to the perception of speech signals, phonetic realizations of a linguistic category (e.g., a phoneme or a word) are extensively variable depending on both linguistically and socially conditioned factors, but listeners promptly access the target representation.

As an example of linguistically-conditioned variation, surface forms of a single phoneme vary dramatically due to influences from adjacent segments, but listeners use coarticulatory information to anticipate successive phonemes in real time (Liberman et al., 1967; Lahiri & Marslen-Wilson, 1991; Dahan et al., 2001b; Beddor et al., 2013). Temporal properties of words are also modulated by linguistic context. Vowel durations of longer words are shorter than those of shorter words and listeners use this correlation to predict the word (Salverda, Danhan, & McQueen, 2003). In addition, as will be reviewed in Section 1.3, words are realized with substantially different phonetic details across talkers, and between-talker variability is often socially conditioned (i.e., can be parameterized by social categories, such as age, gender, regions). Nontheless, even infants show the ability to recognize familiar words, adapting to pronunciations of novel talkers (Swingley, 2005).

The capacity to rapidly adapt to talker variability raised a question dubbed "the lack of invariance problem" (Liberman et al., 1967) that called for decades of research. How are phonetic variants encoded in the mental representation? Approaches to this question can vary depending on assumptions about the format of lexical representations. We explore two different accounts for the nature of lexical representations in the following two subsections (1.2.1 and 1.2.2).

1.2.1. Abstractionist models: an invariable approach to variables

In traditional views on speech perception (including the models discussed in Section 1.1), the description of phonetic categories depended on abstract mental representations (i.e., phonemes). The phonological representation of a lexeme then is stipulated as a sequence of invariant phonemes mute on contextual variability, and access to the underlying representation occurs as a many-to-one mapping process. This sort of abstraction is conceptually useful to explain many behavioral patterns, including categorical perception (Liberman et al., 1957), in which gradient within-category changes in an acoustic dimension do not tend to be detected but the perceptual shift between categories occurs rather abruptly.

The abstractionist account of speech perception seeks to implement a fixed processing route that assesses and transforms the sensory input to match the canonical representation. For example, Gaskell and Marslen-Wilson (1996, 1998) proposed a rule-based inference process where the percept of phonetic variants of the word form are transformed with reference to the phonological context. In such models with invariant representations and invariant processing routes, talker-specific phonetic detail (e.g., allophonic variants) or the sources of phonetic variability (e.g., information about social categories) are treated as random noise information. To account for adaptation to between-talker variability under the abstractionist framework, a simplifying assumption about the nature of the sensory input is necessary; systematic variation of phonetic detail should be filtered out during a normalization process.

For example, vocal tract normalization models posit that listeners use perceptually estimated vocal tract lengths of the talker to resolve ambiguities internal to the spectral qualities of auditory signals (Potter & Steinberg, 1950; Ladefoged and Broadbent, 1957). Specifically, vowels produced with different formant values across talkers (e.g., males versus females) can be categorized as the same vowel phoneme because the auditory representation of vowels is dependent upon the ratio of F0 and F1 (Miller, 1989). Therefore, formant variation across talkers is structured in proportion to each talkers' fundamental frequency (which is conditioned by vocal tract length), and listeners use the ratio as perceptual cues to normalize the vowel sound (Peterson, 1961; Fujisaki and Kawashima, 1968; Slawson, 1968).

1.2.2. Phonetically-rich lexical representations

The abstractionist account has been criticized for its simplifying assumption which "hid subphonemic constraints available in the signal" (Magnuson et al., 2013; 434). In pursuit of a simple module consistently generalizable to the mapping of ever-changing phonetic forms, the influence of subphonemic variation across talkers had been underrated in formal linguistic theories in the last century.

As opposed to categorical perception of gradient acoustic cues (Liberman et al., 1957), there are studies reporting fine-grained sensitivity to sub-phonemic differences (Pisoni & Tash, 1974; McMurray, Tanenhaus, & Aslin, 2002). For example, response times for discrimination of a pair of synthesized tokens on a continuum between /ba/ and /pa/ were conditioned by the size of VOT difference of the stimuli, either when the two stimuli belonged to the same category or when they straddled between categories (Pisoni & Tach, 1974).

The sensitivity to fine-grained phonetic differences is also demonstrated by listners' memory about between-talker variability. There is evidence that memory traces of individuals' voice characteristics are retained in long-term memory (Papcun, Kreiman, & Davis, 1989), and a growing body of literature suggest that lexical processing benefits from talker-indexed phonetic details of an utterance (Mullennix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993; Goldinger, 1996; Creel, Aslin, & Tanenhaus, 2008; Creel & Tumlin, 2011). Specifically, words are recognized more accurately and quickly when spoken by the same individual that the listeners previously heard producing those words during the experiment session, rather than when produced in a novel voice (Mullennix et al., 1989; Palmeri et al., 1993), and listeners can retain talker-specific phonetic detail in the lexical memory for multiple days (Goldinger, 1996). The priming advantage of a previously encountered talker is enhanced when tested with newly learned nonsense words, because the talker-specific realization of the non-word stimuli encountered during the experiment session serves as the only source of memory traces with which listeners can build up the phonetic representation of the nonsense word (Creel et al., 2008).

These findings indicate that knowledge about individuals' difference in phonetic realizations plays a role in lexical-level processing. Conversely, when there is less predictability in talker-specific phonetic patterns (e.g., listening to a talker with greater variability in cue distribution) segment identification responses are slowed down (Newman, 2001). Some scholars interpret the talker-specificity effect as evidence for the existence of multiple phonetically-rich representations for a single lexeme that compete each other for recognition, and argue that experienced talker-specific phonetic detail is encoded in lexical memory (Palmeri, et al., 1993; Goldinger, 1996; Johnson, 2005; see also Nygaard, 2005, for a review of this view).

The view that acoustic memories are stored at word-level storage is also supported by studies on word perception amid environmental noise (Creel, Aslin, & Tanenhaus, 2012; Pufahl & Samuel, 2014), which demonstrate that words are recalled better when listeners reencounter the word in the same noise condition (e.g., dogs barking or bells ringing) that they were previously exposed to along with the lexical items. Taken together, the studies outlined above suggest that representations of previously encountered words contain perceptual links between words and detailed memories of an individual's voice and even implicitly associated environmental acoustics that are not closely tied to the lexicon.

On top of that, as will be demonstrated with empirical evidence in the next section (1.3), talker information provides more robust context-dependent predictability than once thought – especially when it is systematically predicted by socio-indexical categories – and listeners are able to strategically use socially-indexed information during real-time processing of speech input. Thus, the claim for variable and multiple representations of a word is extended in the sociophonetic framework, incorporating individuals' experience with socially-conditioned phonetic variation (see Section 1.4). The detailed mechanisms integrating such lexical properties are implemented in exemplar-based models (see Section 1.5.1), whereas Bayesian models also provide a conceptually different mechanism, supposing selective adaptation of talker-specific access pathways to a single abstract representation (see Section 1.5.2).

1.3. Socially-indexed predictability of phonetic variants

The view of phonetically-rich representations sharply contrasts with the traditional account. Perceptual normalizations – or, more generally, abstractionist mechanisms positing invariable representations and access routes – presuppose a process that warps and neutralizes the raw talker-specific information to match it with the fixed shape of the stored representation, which entails reduction or loss of information (Pisoni, 1997; Johnson, 1997). However, estimation of vocal tract size, after all, is not a consistent and reliable parameter in reality; what listeners actually perceive is not only the concrete cues based on algorithmic sensory information concerning physical properties but also subjective impressions about the *talker*[4] based on individuals' social experiences and expectations (Johnson, 1990; Johnson, Strand, & D'Imperio, 1999).

Escaping from the abstractionism, the sociophonetic framework brought about a surge of interest in (1) how phonetic variants are mapped in accordance with socially-conditioned factors, (2) to what degree socio-indexical information is utilized in linguistic processing, (3) the role of listeners' experience in the formation of long-term representations of phonetic variables and lexical items, and (4) how the interplay between linguistic and social information can be integrated into computational and cognitive mapping. In this section, I discuss research in sociophonetics with these questions in mind.

---

[4] Many researchers in sociophonetics differentiate "talkers" from "speakers". The former means listener's "perceptual impression of a person's identity based on hearing or watching the person speak". In Johnson et al.'s distinction, talker representation is a combination of "objective facts about a person, such as vocal tract length" (1999:365) and socio-cultural factors.

Even a single phonetic category in the same phonetic context is realized with idiosyncratic variation across individual talkers (Newman et al., 2001; Allen, Miller, & DeSteno, 2003). But, talker-specific variability in cue distributions is also conditioned by physical factors, – e.g., vocal tract length (Peterson and Barney, 1952), glottal spreading (Chodroff & Wilsong, 2017) – or by stylistic/context-specific factors – e.g., articulatory habits (Johnson & Beckman, 1996), articulatory effort (Lindblom, 1990), and speech rate (Miller & Liberman, 1979).

Between-talker variability is also systematically structured under macro-sociological categories, such as age, gender, and regional dialect (Labov, 2001). In addition, speakers cultivate and adhere to stylistic forms in line with their communities of practice (Eckert & McConnell-Ginet, 1992) and their roles within their communities of practice (Eckert 2000), while also orienting to the above-mentioned macro-sociological categories (Eckert 2000). For example, speakers' linguistic variation patterns according to their local stances (Labov, 1963), social persona (Eckert, 1989), social prestige (Frazer, 1987), or sexual orientation (Podesva, Roberts, & Campbell-Kibler, 2002). Thus, variant forms are closely indexed to one's identity and social stance (Labov, 2001; Eckert, 2000, 2008).

During perception, listeners attribute social information to the talker based on phonetic cues in the signal. For example, perception of the talker's dialect (Preston, 1993; Clopper & Pisoni, 2004, 2006) or social judgment (Campbell-Kibler, 2007; Levon, 2014) is affected by manipulation of phonetic cues. Further, talker-specific variation in phonetic cue distributions help listeners recognize the talker's sex and identity, even when natural voice quality is removed (Remez, Fellowes, & Rubin, 1997; Fellowes, Remez, & Rubin, 1997).

The reverse is also true; a growing body of literature in experimental sociophonetics has established that, once socio-indexical information is accessed via either auditory or non-linguistic cues, it can affect the perception of phonetic categories. For example, Strand and Johnson (1996) demonstrated that listeners' identification of synthesized fricatives in a continuum between [s] and [ʃ] was affected by auditory and/or visual cues to talker gender. The effect corresponded to the difference in the spectral quality of those segments produced by male and female speakers, i.e. center of gravity (CoG), which is conditioned by the vocal tract size. In addition, the effect was gradient; perceptual boundaries of [s] and [ʃ] for non-prototypically male and female voices were found to exist between the boundaries for prototypical male and female voices. In a follow-up study, Johnson, Strand, & D'Imperio (1999) found that listeners' perceptual boundaries of vowels were affected even when participants listened to a gender-neutral voice and were merely asked to imagine the talker as either male or female.

These findings suggest that phonemic representations are not as static as once thought. Listeners are sensitive to fine-grained socio-indexical properties about the talker, and categorization of speech sounds is adjusted according to the subjective representation of talker characteristics. Stemming from this view, the body of research in socially-conditioned perception of spoken language has been growing steadily over the past two decades. It has been shown that listeners rely on various pieces of socio-indexical cues attributed to the talker and their probabilistic associations with phonetic variants while resolving auditory ambiguity. The utilized talker information include gender (Strand & Johnson, 1996; Johnson et al., 1999), sexual orientation (Munson, Jefferson, & McDonald, 2006), regional dialect (Niedzielski, 1999), ethnicity (Staum-Casasanto, 2008), social class (Hay, Warren, & Drager, 2006b), social persona (D'Onofrio, 2015) and, of particular relevance to the work presented in this dissertation, age (Hay et al., 2006b; Koops, Gentry, & Pantos, 2008; Drager, 2011).

In Drager's (2011) vowel identification test, New Zealand English (NZE) listeners heard resynthesized vowel tokens in a continuum between the TRAP[5] and DRESS vowels in various voices and identified which word they heard. The two vowels are involved in a chain shift in NZE, in which younger speakers' pronunciation of TRAP is raised to overlap with the space of DRESS. Each voice was paired with a photograph of either a younger or an older face. The older participant group (but not the younger group) tended to identify the ambiguous vowels as TRAP when younger faces were presented, showing that the perceptual boundaries of older listeners are biased by the association of younger speakers with a raised vowel space of TRAP.

Additionally, the effect of socio-indexical information on perception of phonological categories can be strengthened by listeners' awareness of linguistic variables (see Drager & Kirtley, 2016, for a review). In Niedzielski (1999), Detroit listeners' perception of the diphthong /aw/ produced by a Detroiter is biased by a stereotype[6] that Canadians produce a raised variant, even when the dialect spoken in Detroit also indeed raises the diphthong. Hay and colleagues (Hay, Nolan, & Drager, 2006a; Hay & Drager, 2010) further show that mere exposure to a social concept involved in a dialectal stereotype (e.g., a written label of the region or a stuffed toy associated with the region) is sufficient to affect vowel perception, even when listeners have no reason to believe that the regionally-associated primes are in any way related to the talker. Thus, indexical properties of phonetic forms are not solely dependent on usage-based distributional causes but can also be reinforced by higher/abstract level links when the variant is socially

---

[5] The lexical sets established by Wells (1982) are used here to refer to the NZE vowel classes.

[6] As opposed to *indicators* and *markers*, stereotypes refer to linguistic variables that exist at the conscious level of listeners' attention to stylistic variation, ones for which individuals in the speech community can discuss the relation between the linguistic variable and its associated social group (Labov, 1972). We follow this definition in this dissertation.

salient[7]. In a similar vein, Sumner et al. (2014) argue that socio-indexical knowledge does not necessarily rely heavily on token frequency because not all phonetic variants are equally meaningful, arguing instead that a socially-idealized variant can be encoded as strongly as a frequent form. They propose a dual-route mechanism for speech processing, in which the lexical-encoding process interacts with socially-weighted encoding.

1.4. Socially-conditioned lexical representations

In this section, we discuss empirical findings which show that lexical processing is influenced by socially-indexed phonetic realizations. The body of sociophonetic perception literature outlined in Section 1.3 demonstrates that there are uneven distributions of phonetic cues across groups of talkers and that these distributions have consequences in sound processing. Also implied in those studies is that the effect of talker-indexed phonetic patterns is not limited to utterances spoken by talkers encountered during experimental sessions, but listeners are also sensitive to the generalized distribution of phonetic variants over social categories. Thus, we might expect that phonetically-detailed memory is not only temporarily stored in the word-level storage as shown in Section 1.2.2 (Mullennix et al., 1989; Palmeri et al., 1993; Goldinger, 1996; Creel et al., 2008; Creel & Tumlin, 2011), but the long-term representations are also shaped by socially-conditioned experiences with the word.

This position can be supported by results from lexical priming experiments using phonetic variables. For example, Sumner and Samuel demonstrate that target word recognition is

---

[7] We return to the issue of multi-layered associations between linguistic categories and social information in Chapter II, where I compare the influences from distributions of word usage and stereotypes about words based on results from Experiment 1.

facilitated when listeners are exposed to a prime word that contains a variant of the listeners' experienced dialect (Sumner & Samuel, 2009) or a variant that is socially idealized (Sumner & Samuel, 2005). In addition, cross-language word processing is also influenced by indexical information of phonetic forms. In Szakay, Babel, and King (2016), bilingual New Zealanders' recognition of L1 lexical targets (English) is primed by translation-equivalent L2 words (Māori) when listening to a talker with an L2 accent, but not when listening to a talker with an L1 accent, indicating that socio-indexical property (i.e., ethnicity) is shared between perceptual representations of an L2 word and an L1 word realized in L2 pronunciation.

There is also evidence that listeners can strategically use the implicit knowledge about socio-indexical links to adapt to the talker's phonetic realization by altering their expectation about what words spoken by a novel talker would sound like. Using a talker of an English dialect where the vowel /æ/ is raised before /g/ but not before /k/, Dahan, Drucker, & Scarborough (2008)'s visual world experiment show that prior exposure to a word that contains the raised variant (e.g., *bag*) leads listeners to adjust their representations in the lexicon in accordance with the vowel space of the dialect, resulting in facilitated disambiguation of a word that would be temporarily ambiguous in other dialects (e.g., *back*).

So far, we have focused on how representations and lexical access can be affected by socially-conditioned talker-indexed memories, but some proponents for usage-based grammar (e.g., Bybee, 2001; Pierrehumbert, 2001, 2002) claim that word representations are encoded with linguistically-conditioned word-specific phonetic memories. In this view, phonetic realization of words is conditioned by word-specific contextual factors (e.g., token frequency, neighborhood density, semantic categories). First, repetition of lexical use leads to reduction of the phonetic form through entrenchment of patterns to fulfil discourse-oriented functions (Haiman, 1994;

Bybee, 2001). Consequently, lenitive changes are often led by realizations of high-frequency words (Bybee, 2001; Pirrehumbert, 2001; Hay & Foulkes, 2016). For example, reduction or deletion of /t/ and /d/ in English occurs more frequently for words with higher lexical frequency (Hooper & Bybee, 1976; Bybee, 1985; 2000; 2001; 2002; Gahl, 2008).

In contrast, as an example of fortition, vowels tend to be hyperarticulated when producing words with high neighborhood density (i.e., words that have many neighbor words that differ in just one phoneme) to reduce phonological competition (Wright, 2004). Additionally, vowel production in Montreal French is conditioned by lexical classes (e.g., semantic categories) (Yaegor-Dror & Kemp, 1992).

One pivotal factor for lexical processing this dissertation illuminates is grounded on a view that socially-conditioned phonetic memories are accumulated in word-level storage (as well as in phonemes), and so production and perception of words are modulated by associations between words and their social and linguistic context (Hay, forthcoming). In this view, word-specific phonetic features are not just linguistically constrained but also socially conditioned because certain words are more frequently used by some people than others. The skewed distribution of word usage has its consequence on phonetic detail encoded in the representations; since usage frequency and phonetic properties of a word often covary across speaker groups, representations of words associated with a social group may be closely linked to characteristic phonetic realizations of those people. For example, Hay and Foulkes (2016) demonstrate that words that are more frequently produced by younger NZE speakers tend to be realized with a lenited intervocalic /t/ (i.e., [d/ɾ]), which is an innovative phonetic variant in ongoing sound change led by younger NZE speakers. This trend was observed even in speech produced by older speakers, in support of the claim that word-level representations are shaped by individuals' past experience with a given word.

When it comes to perception, retrieving phonetic properties indexed to a particular speaker group would activate lexemes that are probabilistically associated with that speaker group. Walker and Hay's (2011) lexical decision test provides evidence that lexical access is improved when the word is produced by a talker from the age group who uses the given word most frequently. To estimate the distributional associations between a word and age, they used two different NZE spoken corpora recorded in different times. Words that appear relatively frequently in one corpus compared to the other were selected as *young words* and *old words*, respectively; words were not "stereotypically" associated with any age group. Therefore, their results demonstrate that recognition facility is affected by the words' distributional properties as determined by relative frequency of occurrences across age groups. They argue that lexical representation is shaped by a lifetime of exposure to the statistical distribution of phonetic realizations, and recognition is improved when the incoming signal resembles the generalized phonetic properties of the social group who produces the word frequently.

Taken together, research findings outlined in this section point to some attributes of lexical representations that had not been captured in the abstractionist approach. Phonetic representations of a word are malleable (rather than fixed), encoded with word-specific contextual factors including socially-conditioned phonetic details, and shaped by individuals' prior experience with the word.

1.5. Cognitive models integrating socially-conditioned experiences

As reviewed in Sections 1.3 and 1.4, language variation and its processing take place largely in principled and predictable ways – especially when social factors are taken into account in addition to language-internal factors – and listeners cope with the variable nature of acoustic signals by integrating probabilistically meaningful characteristics that arise from talker-specific and word-specific phonetic patterns into speech processing routines. This observation demonstrates the need for an experience-based model, in which our cognitive mechanism makes use of prior experience with language variation as a rich source of critical information.

The basic assumption of experience-based models is that listeners accumulate implicit knowledge about statistical distributions of variant forms over the lifetime. Through this process, talker-specific fine details of phonetic realizations and their indexical properties are integrated into the cognitive system for the efficiency of perception (Pisoni, 1997). Thus, either phonetic representations or the pathways to access the representations must be flexibly updated by linguistic and social change in a consolidated framework, increasing adaptability to the covariance between phonetic forms and speakers.

1.5.1. Exemplar-based approach

Exemplar models of human perception (Nosofsky, 1988, 1991; Estes, 1993) are adopted to the study of speech perception. In exemplar models of speech perception (Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001, 2002), experiences with phonetic realizations of a lexical item are individually registered in episodic memory as phonetically-rich exemplars and encoded within the mental representation[8]. Thus, a speech category (e.g., phonemes, words) is composed of a set of experienced instances of the category (Pierrehumbert, 2002), and perceiving an utterance involves mapping the incoming acoustic signal to an existing exemplar category based on perceptual similarity (Johnson, 1997).

Exemplar models provide an efficient mechanism for adaptation to talker variability. Reencountering a talker can activate the stored exemplar of the known talker and result in a processing advantage (Mullennix et al., 1989; Palmeri et al., 1993; Goldinger, 1996). When listening to a novel talker, talker-appropriate exemplars are activated while non-relevant exemplars are deactivated (Nosofsky, 1988). In this process, the activated set of exemplars are selectively tuned to any kinds of relevant talker cues, so that linguistic representations are adjusted according to the perceived identity of the talker (Johnson, 2005).

Through continuous exposure to phonetic variability, exemplar clusters are accumulated forming multi-modal distributions according to the socially-structured phonetic patterns. Then, awareness of socio-indexical properties emerges from experience with the statistical distributions of phonetic forms over social categories, and these social indices are acquired in an early age (Foulkes & Docherty, 2006). Thus, phonetic exemplars are indexed to relevant social information about the talker (e.g., gender, age) (Foulkes & Docherty, 2006; Johnson, 2006). During speech perception, such higher-level, abstract social information indexed to the exemplar can be activated either by acoustic or non-linguistic cues, and the activation can then spread to talker-appropriate exemplars of phonetic variants (Hay et al., 2006a, 2006b; Hay & Drager,

---

[8] These exemplar-based representations can be at any level of linguistic analyses, including segments, words, and even morpho-syntactic level (Bybee & Cacoullos, 2008). However, lexical representations are the main interest of this dissertation.

2010). However, activation of phonetic exemplars can also directly activate phonetic variants or lexemes that are associated with the same social category, even without awareness of the talker identity[9].

The framework using episodic memories emphasizes the flexibility of a lexicon, positing that representations are susceptible to, and are accessed by means of, word-specific memories of systematic phonetic patterns. The plasticity of memories stored in the lexicon provides an elegant theoretical ground for listeners' ability to utilize the accumulated experiences with phonetic realizations of words over social categories (Sumner & Samuel, 2009; Dahan, et al., 2008; Walker & Hay, 2011). However, not every single exemplar with allophonic variation can be retained in the long-term memory; episodic memories involved in linguistic processing, just like non-linguistic memories, decay over time if not activated frequently. Thus, an efficient mechanism should also be able to generalize frequently occurring patterns of word-specific properties as fixed properties of the representation, enabling fast adaptation to novel word forms or novel situation (Pierrehumbert, 2016).

This sort of generalization processes is best understood in a hybrid account (e.g., Pierrehumbert, 2002, 2003, 2016; Pitt, 2009; Pinnow & Connine, 2014), which combines strengths of abstractionist models and episodic models. The hybrid models posit variability in representations and also posit a higher-level abstract representation for a given word at the same time (being underspecified for certain detailed phonetic features). Abstract-representational

---

[9] In Chapter III, we will discuss whether activation of social information can be abstracted away in this process based on the results from Experiment 3. This issue is also related to the question addressed in Experiment 2, whether socio-indexical information is encoded only as top-down contextual cues influencing prior expectancy, or whether the signal is directly compared to acoustically-detailed representation. We will come back to a general discussion of this issue in Chapter V.

features are weighted by frequency of each variant, allowing room for generalization over indexical properties. In such a framework, an individual improves recognition of variant forms by developing multiple levels of representations during their lifetime, and the representations are generalized and refined based on the statistical regularities found through additional experience (Pierrehumbert, 2003). Through this generalization process, frequently activated exemplars are associated with a variety of episodic memories, forming dense exemplar clouds, which results in a processing advantage (Pierrehumbert, 2001). Exemplars stored in long-term memory then can either establish a representation that is closely associated with an existing category, or be generalized into an existing category if the incoming speech signal is sufficiently similar in its phonetic characteristics (Pierrehumbert, 2001; Bybee & Cacoullos 2008).

1.5.2. Bayesian approach

From a Bayesian perspective, human perception is an unconscious inference process where uncertainty of the sensory input is resolved by probabilistic integration of prior knowledge. When Bayesian inference is adopted to speech perception, listeners approach the optimal representational node by concurrently combining the perceptual evidence with knowledge of the prior probabilities of words (Norris, 2006; Norris & McQueen, 2008). In such processes, talker information in the unfolding signal becomes one of the most crucial factors for the probability (Kleinschmidt & Jaeger, 2015).

The fundamental assumption of the Bayesian account differs from connectionist models (e.g., exemplar models, TRACE, Shortlist), primarily in that they do not posit the interactive-activation network. An alternative interpretation of influences from contextual information is put forward with computational algorithms using Bayes’ theorem in a unidirectional feed-forward system (i.e., no similarity-based feedback from high-level nodes), under the assumption that what is attributable to top-down activation in the connectionist account is indeed fully explicable by jointly adapted bottom-up functions in a rule-based architecture (Norris & McQueen, 2008: 383)[10].

Unlike traditional rule-based models, however, Bayesian models can handle speech variability in real time by allowing flexibility in access routes. When the signal is completely unambiguous or when the listener is probabilistically familiar with the situation (e.g., a predicted word and/or a known talker), the signal is easily mapped to the best-matching word. On the other hand, under substantial uncertainty of the acoustic input (e.g., a surprising word and/or a novel talker), the model approximates the optimal decision of the speech signal by using a Bayesian strategy, given the constraints imposed by the speech signal and potentially all kinds of contextual knowledge that may influence the prior probability of the words. As uncertainty increases, the influence of prior probabilities (i.e., an influence of readily known contextual information) increases, transforming prior expectancy about what the talker is going to say.

The mechanism for real-time integration of socio-indexical information is specifically defined in Kleinschmidt & Jaeger (2015). Their *ideal adaptor* framework proposes a belief-updating learning system, in which listeners' beliefs about the talker are simultaneously updated by the unfolding cue distributions, and the generative model is perceptually recalibrated and

---

[10] Connectionism takes a holistic approach to human cognition by supposing an activation network system of stored memories and detailed representations, while specific details of neural functioning and sensory algorithms tend to be abstracted away. In contrast, Bayesian models are dedicated to computational mechanisms, focusing on the input-output relations needed for decision-making processes during access to abstracted representations. Whether the two are mutually exclusive or are useful to describe co-operation between different levels of the entire system is a pending question in the cognitive science (Jurafsky, 2003).

selectively adapted in accordance with the situation, balancing between stability and plasticity of the adaptation depending on how familiar the situation is.

Because prior probability plays a crucial role (instead of activation) in this completely bottom-up mechanism, rapid and automatic processing of social information is obligatory, guiding access to the most plausible interpretation of the signal with respect to concurrently updated knowledge about covariation between social characteristics and phonetic cues. Studies using ERPs suggest that recognition of social information is encoded as a routine process that automatically occurs simultaneously with linguistic inference. Van Berkum and colleagues (Van Berkum et al., 2008; Tesink et al., 2009) report that extraction of social information from the speech signal takes place in a shared brain region with the decoding of linguistic meaning of the word, as early as 200-300ms after the onset of a spoken word.

1.6. Dissertation overview

As a summary of the review so far, the aforementioned work in psycholinguistics presents different predictions for the time course of top-down influence of lexical information (Section 1.1) and for the degree to which phonetic-level detail is abstracted for lexical processing (1.2). Research in sociophonetics provides evidence that phonetic processing is influenced by listeners' implicit knowledge about the relationship between the talker and cue distributions (1.3). Empirical findings on the lexical-level processing provide converging evidence that listeners use socially-indexed phonetic cues during lexical recognition (1.4). Finally, predictions of experienced-based models also converge on a consensus that word-level representations or access processes are tuned to listeners' sociolinguistic experiences with the word (1.5).

Of particular relevance to the role of word-level associations of social information, Walker and Hay (2011) show that lexical access is facilitated when the word is produced by a talker from an age group that the word is associated with. They interpret that the processing advantage of age-congruent voice is led by an acoustic or indexical match between the signal and the representation (rather than by contextual priming of age-related talker information), because their analysis demonstrates that the effect is parameterized by purely probabilistic associations (i.e., relative token frequency of words), but not by associations generalized into listeners' conscious age-related bias for the word (i.e., word-stereotype). Building on their work, this dissertation further examines the interplay between socially-indexed phonetic forms and lexemes. Results from four experiments are reported to support and extend their claim by demonstrating that age-indexed phonetic forms can rapidly and directly index their associated lexemes with little resort to activation of abstract social information, and that the effect can also arise for words that are stereotypically associated with age. In addition, the temporal aspects of age-indexed phonetic processing will be discussed based on the results. Specific agenda are summarized below.

Chapter II will further examine the underlying principles of the age-congruence effect using two modified versions of Walker and Hay's (2011) lexical decision experiments. Experiment 1 explores the possibility that probability-based indices between lexical items and social information are refined by socially-salient indexical information of words, so that lexical processing is also affected by stereotypical associations between words and social groups – as well as distributional factors – when appropriate lexical items are used. Experiment 2 will test whether the effect of age-indexed phonetic cues can arise rapidly even when no prior cues about the talker are available before hearing the word's onset. If an interaction between the talker's age and the word age is observed in both experiments in Chapter II, it would expand our understanding about the indexical information of words, and lend support to Walker and Hay's assertion that phonetic detail is encoded in the lexicon.

In Chapter III, I present results from a third lexical decision task that is designed to test whether an age-congruence effect can be observed when the phonetic cues to age are limited to a single phonetic variable contained in a word that precedes the lexical target (i.e., a prime word). An effect was predicted because exposure to a single sociophonetic variant would either (a) trigger age-related expectations about the talker, and so the activation spreads to associated lexemes, or (b) directly index age-associated lexemes. Accordingly, also discussed is whether activation of age-related social information is necessary to observe an age-congruence effect. If a direct influence is found, it would provide further evidence that phonetic variants and lexical items are closely linked in cognition.

In Chapter IV, a lexical identification task is presented to test whether listeners use sociophonetic cues in the signal to anticipate what word they will hear, when information about the lexical items and the talker is provided prior to the auditory target. This time, I use an eye-tracking method to explore the time course of real-time integration of unfolding phonetic cues. The results are discussed focusing on the time course of the age-indexed processing and listeners' strategic use of sociophonetic cues in perceiving socially-incongruent phonetic realizations.

Finally, Chapter V will conclude the dissertation by presenting a general discussion about the overall implications and suggestions for broader questions that need to be addressed. Integrating results from the experiments, the discussion will focus on (1) the degree to which phonetic realizations and words are linked in the underlying mechanism of speech processing,

(2) the role of salience of social information in phonetic variants and words, and (3) how existing perception models can be developed to account for these aspects of lexical processing.

Additionally, that the vast majority of the work is based on English is problematic for language perception models because an influence of sociophonetic variation may be manifested differently in a different social context. Investigation with Korean, a language that is genetically unrelated to English and interacts with its own social constructs, would help demonstrate that the effect of sociophonetic variation on lexical processing is widespread.

CHAPTER II

Age Associations and Lexical Access

As reviewed in Chapter I, cognitive associations between linguistic variants and social knowledge are formed and shaped by individuals' experience with socially-conditioned phonetic variation. The role that the associations play in language processing is evidenced by the findings that speech recognition is influenced by socio-indexical information attributed to the talker. Particularly as an influence on lexical access, Walker and Hay (2011) demonstrate that English listeners recognize words faster and more accurately when the talker's age is congruent with the age of the speakers who produce the words frequently. In Chapter II, I report results from two experiments that replicate their finding using lexical items in Korean and discuss the implications in light of speech perception models.

2.1. Experiment 1: Effects of age-related distribution and stereotypes

In Walker and Hay (2011), the age-congruence effect is found when the association between a word and age (*word age*, henceforth) is determined by relative frequency between words produced by older and younger speakers over two corpora collected in different time periods. As post hoc analysis, they also tested whether the effect is predicted by word ages based on native speakers' beliefs about whether each word is more likely used by younger or older people. Since they found no evidence that lexical access was influenced by conscious awareness of word distributions, the age-congruence effect is interpreted as a consequence of listeners' lifetime exposure to the skewed distribution of word use, which is consistent with the predictions of experience-based cognitive models that human cognition is largely parameterized by probabilistic factors.

However, while actual experience with people from various social groups is crucial in the formation of cognitive links between phonetic and social information, the awareness of sociolinguistic variation can also be enhanced by generalization processes at an abstract level, such as metalinguistic discussions about social groups and the way they talk (Labov, 1972). This raises the question: Can talker age also influence access of words that are *stereotypically* associated with speakers of different ages and, if so, is there a difference in effect size?

Experiment 1 explores the question using a modified version of Walker and Hay's (2011) experiment, with lexical stimuli selected from a wide spectrum of word-stereotypes in Korean. Word age is treated in two folds: usage age and stereotypical word age. *Usage age* is determined based on a cross-age comparison of self-reported measures of how frequently they verbally produce each word. On the other hand, *stereotypical word age* is measured based on native speakers' judgment on whether younger or older people are more likely to use each of the words. The methods of calculating each of these measures are described in Section 2.1.1.1. The two types of word age are used as predictors for recognition facility and their effects are compared to examine whether lexical access benefits more from distribution or stereotype.

I hypothesize that recognition of age-associated words will be faster and more accurate when the talker's age is congruent with word age (either usage- or stereotype-based), and that the effect of stereotypes will be greater than that of frequency, suggesting that distributional associations are reinforced by social stereotypes.

2.1.1. Method

2.1.1.1. Word age rating. An online survey was conducted to choose lexical items used in the experiment and assign the two types of word ages to each item. Survey respondents were 80 Seoul Dialect speakers living in the Seoul Metropolitan area, a separate population from the lexical decision experiment participants (see Section 2.1.1.5 for experiment participants). There were 42 older (aged from 50 to 71) and 38 younger respondents (college students aged between 18 and 25). 340 words (2-4 syllables) were included as survey items, which I selected from three potential word age categories: 116 young, 179 old, and 45 age-neutral words. The respondents' task was to provide self-reported measures for their own exposure level, frequency of verbal use, and age-related stereotype for each item.

Young word candidates were chosen from three lexical categories: (1) words that are semantically associated with campus life (e.g., *kkwacampa*[11], 'a jacket worn by students in a department to mark their sense of solidarity'[12]), (2) coined words that are widely used in computer-mediated communication (e.g., *notap*, 'no solution', a word made of *no* in English and *tap*, 'a solution' in Korean, sarcastically describing someone's poor performance or stupid behavior), and (3) words for new concepts (e.g., *pheyisupwuk*, 'Facebook').

---

[11] Phonemic transcriptions in this dissertation are based on "The Yale Romanization System".

[12] Definitions of the young word examples are based on an online dictionary of Korean neologism, Naver Open Dictionary (http://kin.naver.com/opendic/index.nhn), except kkwacampa, which was not found in any of the available dictionaries. Definitions of old word examples are based on the ET-house Neungyule Korean-English Dictionary (http://www.et-house.com/pages/dictionary/ko_dic.asp) and the Standard Korean Dictionary of National Institute of Korean Language (http://stdweb2.korean.go.kr/main.jsp).

Old word candidates consisted of four categories: (1) words that are related to rural life (e.g., *cangtoktay*, ‘a platform (in a traditional house) used to store crocks of sauces and condiments’), (2) vanishing terms (e.g., *kwukminhakkyo,* an antiquated term for ‘an elementary school’), (3) kinship addressing terms that had been frequently used under an extended-family system (e.g., *ansaton*, ‘one’s daughter-in-law’s [or son-in-law’s] mother’), and (4) loanwords that had been imported from Chinese and Japanese before the independence from the colonization by Japan in 1945, and then became discouraged by the so-called “National Language Purification Policy” (e.g., *syassu*, ‘a shirt’, *pwullanse* [佛蘭西], ‘France’).

Since the participants in the experiment (see Section 2.1.1.5) are all young college students and they are expected to be familiar with young words but not with old words, it is possible that lexical familiarity (or frequency) is correlated with age-related information. To alleviate the effect of this association, the experiment stimuli included two groups of lexical items that are not apparently associated with either age group (and do not belong to any of the potentially-aged categories), but vary in terms of frequency. I refer to one of them as fillers (N=96), which were selected from the most frequent words in the Sejong spoken corpus, and the other is neutral words, which are ones that are not frequently found in the corpus. 45 neutral word candidates were included in the survey items (e.g., *kkatalk*, ‘reason’, *yutaykam*, ‘a sense of fellowship’, *sonkalakcil*, ‘finger-pointing’), but fillers were not (and therefore were not assigned word age) due to a large number of survey items.

The 340 words were divided into two sets of 170 words. 40 respondents (7 older males (OM), 14 older females (OF), 11 younger males (YM), 8 younger females (YF)) answered questions for one set (58 young, 90 old, 22 neutral words), and the other 40 respondents (8 OM, 13 OF, 7 YM, 12 YF) were given the other set (58 young, 89 old, 23 neutral words). Each word was presented in its written form and accompanied by three questions. Items with non-standard phonetic forms were presented in their surface forms, rather than their standard orthography (e.g., 빠다 (*ppata*), instead of 버터 (*pethe*), 'butter').

The first question asked how often the respondents see or hear other people using each word to measure word familiarities. Respondents chose their answer from four statements about the degree of exposure, and their answers were coded as *reported-exposure scores*, from 0 ("I have never seen or heard the word, and I don't even know the word.") to 3 ("I see or hear it pretty often, at least once a week or more frequently.").

The second question asked how often they use the word in real conversation to measure frequencies of verbal use. Respondents' answers were coded as *reported-usage scores*, from 0 ("I have never spoken the word.") to 3 ("I use the word pretty often, at least once a week or more frequently.").

The third question was asked only of those who did not choose 0 in the first question (i.e., excluding those who do not know the word): "Between younger people (10s-20s) and older people (60s and above), who do you think uses the word more frequently?" Respondents chose from five statements, and *stereotype ratings* were coded from -2 ("Younger speakers use it much more frequently."), through 0 ("Both use it about the same amount."), to +2 ("Older speakers use it much more frequently.").

Two predictors of word age were drawn for each of the 340 words, based on the ratings for the last two questions. First, *stereotype score* (ST score) was obtained by calculating the mean of the stereotype ratings for each word that only the young survey respondents provided. Only responses from the young respondents were used because they are in the same age group as participants tested for the experiments reported in this dissertation. Second, the *usage-age score* (UA score) was calculated for each word by subtracting the mean of the younger respondents' reported-usage scores (in the second question) from the mean of the older respondents' reported-usage scores. Therefore, a high (positive) UA score indicates that the word is used more by older people, and a word with a low (negative) score is used more by younger people.

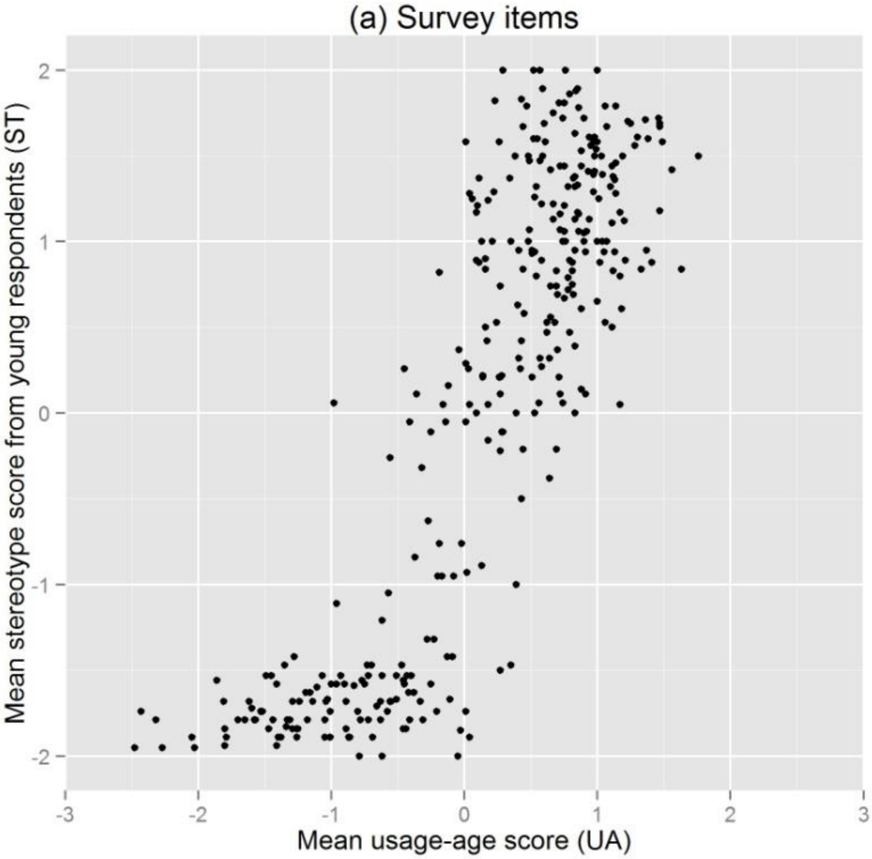
The ST scores do not perfectly correspond to the linguistic stereotypes in the sense of Labov's (1972) definition in that they were not extracted from native speakers' "metalinguistic commentaries". Nevertheless, since the question and the statements derived the respondents' metalinguistic thinking about word usage difference across age, the ratings represent their conscious beliefs about the associations between words and age, which can be dissociated from the ratings for their own actual usage frequency.

It is also noteworthy that the UA scores were obtained from self-reported frequency, and that the survey respondents' rating on frequency may have been affected by the stereotypes that were also asked about and vice versa. This contrasts with Walker and Hay's (2011) measure of token frequency, which was corpus-based. This could not be done because there is no available corpus in which the items I used (particularly young-associated words[13]) are found and coded for speaker age. However, introspective measures of lexical knowledge, such as self-reported measures of familiarity and frequency, have been attested to be an efficient method to capture the characteristics of word representations shaped by individual's experience (Kuperman & Van Dyke 2013). Thus, despite being a subjective parameter, the UA score likely portrays an accurate relative representation of word frequency based on individuals' representational differences.

---

[13] Most of the young-associated items are non-standard words that are not found in dictionaries. They appear only in some online dictionaries of neologism, such as the Naver Open Dictionary (http://kin.naver.com/opendic/index.nhn). In contrast, Walker and Hay's (2011) stimuli are all standard English words that appear in dictionaries.

2.1.1.2. Lexical stimuli. From the 340 survey items, 288 words were selected as critical items for the experiment. Plotted in Figure 2.1 is distribution of ST and UA scores for the survey items (a) and the critical items (b).



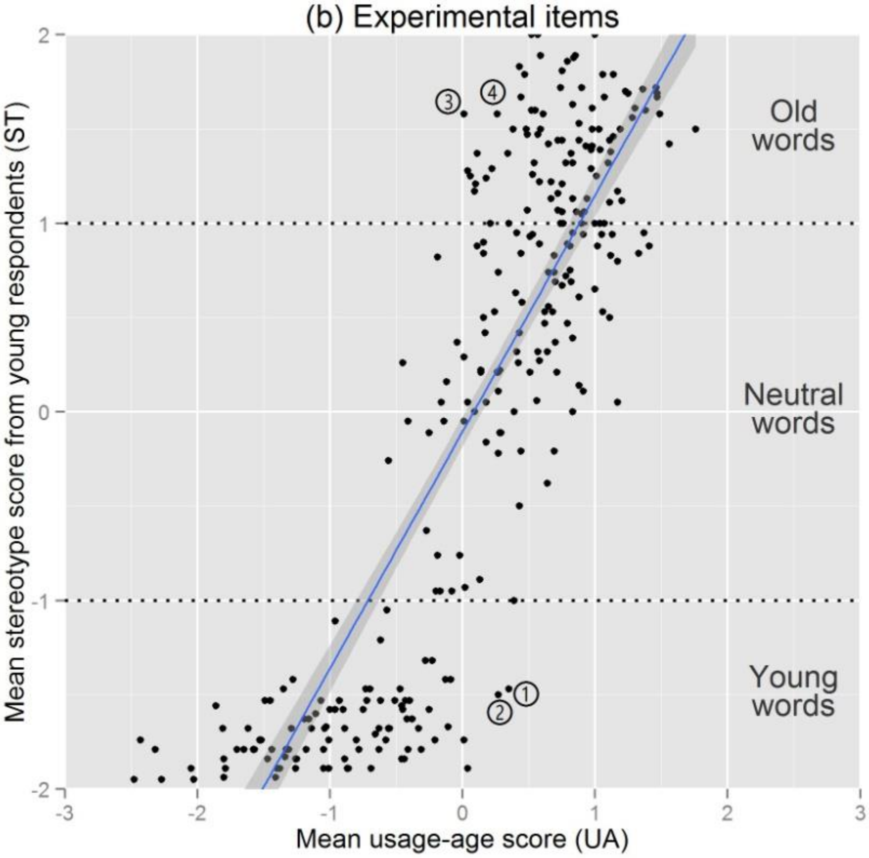


Figure 2.1. Distribution of stereotypical word age (ST) and usage age (UA) scores (for (a) the survey items and (b) the critical items): In (b), circled numbers indicate examples of items with non-matching UA and ST scores. The two dotted lines are boundaries of word age categories, and the solid line represents a linear regression in which ST is regressed by UA. Grey areas indicate standard errors.

Critical items were chosen from three word age categories based on ST score (N=96 for each category). Old words were selected from the words at or above the ST score of +1, young words at or below -1, and neutral words from between -1 and +1. I excluded 52 items based on two criteria. First, nine words were excluded due to low mean reported-exposure scores for the young survey respondent group (under 0.47/3.00) in order to prevent high error rates for the young participants in the lexical decision experiment. Second, 41 items were removed from densely populated regions in Figure 2.1(a), so that items do not cluster around a particular range of ST and UA scores.

Unsurprisingly, the ST scores and UA scores are highly correlated in a Kendall's tau correlation test ($\tau = .71$, $p<.001$). However, there are some noticeable trends in the distribution of data points[14]. Young words are distributed closer to the floors of the scale, i.e., -2 in ST and -3 in UA, than old words are. Neutral words are centered around a region with relatively high ST and UA score, i.e. not around (0, 0). Importantly, there are some words that have low ST scores (i.e. associated with young people), but relatively high UA scores (i.e. used more frequently by older people). These include *khaphwuchino*, 'cappuccino' and *tikha*, 'a digital camera', marked by ① and ②, respectively in Figure 2.1(b). Conversely, some words have high ST scores but more neutral UA scores (e.g., *ipoke*, 'hey' and *tapang*, 'a tea house', marked by ③ and ④, respectively). Examples of words with closely linked ST and UA score are a young word, *notap*, 'no solution', and an old word, *kwucwa*, a word imported from Japanese, meaning 'a bank account'.

In addition to the 288 critical items, the 96 filler items (see Section 2.1.1.1) were also included as real word items, and 384 non-words were created to make up 768 stimuli in total, as summarized in Table 2.1. A list of all real-word items can be found in Appendix I.

---

[14] Note that this data is based on means of ordinal data and are not normally distributed.

Table 2.1. Summary of stimuli

Number of syllables			Two	Three	Four	Total
Real words	Critical	Young	48	36	12	96
		Old	48	36	12	96
		Neutral	35	41	20	96
	Filler		85	10	1	96
Non-words			216	123	45	384

Non-word items were intended to be as word-like as possible to yield sufficient processing time to observe the effects. They were created by changing one or two segments at varying positions of existing Korean words that were not used as word stimuli[15]. Base words were chosen from the same word categories used for potentially old, young, and neutral words. All non-word stimuli are phonotactically legal but do not appear in the Standard Korean Dictionary of National Institute of Korean Language (http://stdweb2.korean.go.kr/main.jsp).

2.1.1.3. Talkers. In order to minimize effects of socio-indexical information contained in the voice other than perceived voice age, the four talkers used in the experiment[16] were selected through a norming test. First, I collected voice samples of 26 speakers recruited from the Korean community in Hawaiʻi (7 OM, 7 OF, 6 YM, and 6 YF). Then, social characteristics of each voice were rated by seven raters who were naive to the research purpose (two males and five females, aged between 31 and 37). Raters heard recordings of six lexical stimuli (two words in each word

---

[15] For both non-words and real words, a large portion of the items were phonetically ambiguous with at least one real-word cohort competitor before the ultimate syllable was heard. Thus, it is expected that listeners do not rely heavily on expectations formed prior to the endpoint of the signal.

[16] I use the term, *talkers*, to refer to the four individuals whose voice was used in the experiment, as opposed to *speakers*, which I use to refer to the general population who speak Korean.

age category) and were asked about age, dialect, socio-economic status, education level, and how likely each speaker is to use language associated with younger people.

Based on the results, four talkers were selected (OM: 60 years old / rated as 53.14, OF: 75 / 63.86, YM: 26 / 33.14, YF: 22 / 20.71), whose ratings most closely met five criteria in their respective age group: someone who is (1) a fluent Seoul Dialect speaker; (2) in middle or upper-middle socio-economic status; (3) between 18 and 25 years old for younger talkers, or over 50 for older talkers; (4) currently a college student for younger talkers, or at least a high school graduate for older talkers; and (5) likely to use young language for younger talkers, or unlikely to do so for older talkers.

2.1.1.4. Recording. Stimuli were recorded by each talker in a sound-attenuated booth at the University of Hawaiʻi at Mānoa. A portable Tascam DR-7 recorder was used with a mono, 32-bit, 44,100 Hz sampling rate setting. Each word was produced twice, preceded by a carrier phrase, *ipen tanenun*, “This word is...”, and the lexical items were extracted from the carrier phrase using Praat (version 5.4.04, retrieved 29 January 2015 from http://www.praat.org/).

Due to the lack of lexical stress in Seoul Korean, it was necessary to control for prosodic differences across and within talkers. All bisyllabic items were produced in a high tone on the first syllable (H+), followed by a low IP-final boundary tone (L%). Tri- and quadri-syllabic items were produced in either of two types of tonal frames, following Jun’s (2011) Revised Intonation Model of Seoul Korean. For words that begin with an aspirate or tense consonant, or /h, s/, the first two syllables were produced in high tones (H+H), and a low IP-final boundary tone (L%) was assigned starting from the third syllable. Otherwise, the first two syllables were produced in a low and a high tone (L+H), followed by L% (see Figure 2.2).

During the recording session, I played pre-recorded samples of my voice to the talkers, so that they could imitate the rhythm and pitch of the samples as closely as possible. I corrected the talkers immediately when mispronunciation occurred.

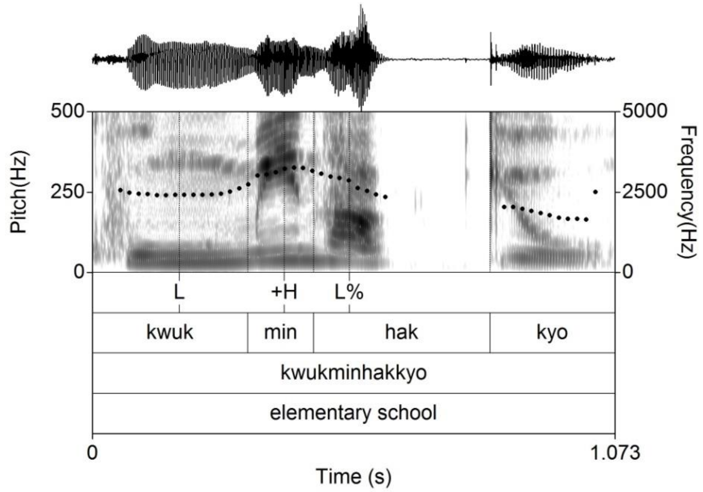


Figure 2.2. Recording example: a quadrisyllabic word, *kwukminhakkyo*, ‘an elementary school’, spoken by the YF talker

2.1.1.5. Participants. 48 native speakers of Korean (34 females and 14 males, aged from 18 to 26) were recruited from two locations: 24 participants from a pool of visiting students at the University of Hawai‘i, and 24 from Chung-Ang University, Seoul, Korea. All the 48 participants were registered students at colleges in the Seoul Metropolitan area and all the Hawai‘i participants had stayed in Hawai‘i less than three months at the time of participation. Data in the item lists 1-4 were gathered from the Hawai‘i participants in January-March, 2015, and then the Seoul participants were tested with the item lists 5-8 in August-September, 2016. No differences in accuracy or reaction times were hypothesized or observed between locations.

All but five participants listed the Seoul Dialect as their most frequently used dialect. The five participants (3 males and 2 females) listed the Kyeongsang Dialect as primary and the Seoul Dialect as secondary, but they all reported they are also fluent in the Seoul Dialect. Participants were paid for their participation.

2.1.1.6. Procedure. The experiment was implemented in E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and was run on four different computers with Windows (two desktops in Hawaiʻi and two laptops in Seoul). In each trial, participants were played an auditory stimulus and pressed one of two buttons on a response pad (model: Cedrus RB-530, or RB-834) to indicate the lexical status of the stimulus: a red button with their dominant hand if they thought they heard an existing Korean word, or a blue button with the recessive hand if not. The two buttons were placed at the far ends of a Cedrus response box. All stimuli were played in isolation in order to prevent any sentential or pragmatic influence.

Participants were instructed to respond as quickly and accurately as possible and not to move their hands off the response box while the experiment is in progress. They were also informed that "a word" in this experiment referred to any variety of word forms that Korean speakers use, by showing various examples of loanwords, non-standard phonetic variations in dialects, and coined words. In practice trials, feedback was provided on the computer monitor screen to inform participants whether their response was correct. Throughout the experiment session, the screen only showed “Experiment in progress” on a white background. But, when participants took longer than 1,300ms to respond, a warning message appeared on the monitor, asking them to respond more quickly.

It took about 45 minutes on average to complete the experiment, which was composed of four blocks (see Section 2.1.1.7). Participants were able to choose to take a short break between blocks, and there was a mandatory break after the second block, during which a 2-minute video presented natural scenery from Hawaiʻi. The experiment session was followed by an exit survey, where participants heard two quadrisyllabic word stimuli in each of the four talkers’ voices and rated perceived social information about each talker (e.g., voice age, socio-economic status, and education level).

2.1.1.7. Design. The experiment session was divided into four blocks with one talker per block. Stimuli were counterbalanced by talker and presentation order of talker age (whether the first two blocks were older or younger talkers). The 48 participants were randomly assigned to one of these eight (4 talkers x 2 presentation orders) lists of stimuli.

2.1.2. Results

The data obtained from all experiments reported in this dissertation were analyzed in R (version 3.4.3, retrieved 1 February 2018 from https://cran.r-project.org/). In this chapter, following the traditional method used in psycholinguistics, frequentist mixed effects models were fit to accuracy (binomial regression) and reaction time (linear regression) using the lme4 package (Bates et al., 2015). Additionally, I ran Bayesian hierarchical models using stan_glmer and stan_lmer function in the R package *rstanarm*, (1) when the model failed to converge when the maximal random effects structure justified by the design (Barr et al., 2013) is included (accuracy analysis in Experiment 1 (Section 2.1.2.4) and Experiment 2 (2.2.2.4)), or (2) when only a marginal effect is found from the frequentist statistical inference (reaction time analysis in Experiment 2 (2.2.2.4)). In such cases, I report results from both methods in parallel, in line with Frank, Trompenaars, and Vasishth (2016).

In this section, I begin by discussing the results for accuracy in Experiment 1 (2.1.2.1), presenting two frequentist models – one with stereotype word age and the other with usage age – in order to demonstrate that both predict accuracy. I then present the results for reaction times (2.1.2.2). Next, I conduct an analysis that compares UA to ST in order to determine whether usage or stereotypes better predicts the links between words and age (2.1.2.3). In 2.1.2.4, I reexamine the accuracy data in Experiment 1 by fitting a Bayesian model. This is worthwhile because Bayesian models can run without a converge error even when the maximal structure is included, and because they provide more direct evidence for the hypothesis (see 2.1.2.4 for detail). Then, in 2.1.2.5, I provide a brief summary of the results from Experiment 1 before presenting Experiment 2 in Section 2.2.

2.1.2.1. Accuracy. Among the 36,864 tokens gathered from 768 trials by 48 participants, there were 31,941 correct responses. The overall accuracy rate was 86.98% for real words and 86.31% for non-words. The low accuracy rates for a lexical decision test are not surprising because the young participants were expected to be unfamiliar with infrequent lexical items (especially old words) and because non-word items were created to be highly word-like. In Figure 2.3, accuracy rates from the real word targets excluding filler trials (N=13,824) are plotted by talker age (old vs. young) and word age (ST score in (a) and UA score in (b), treated as continuous).

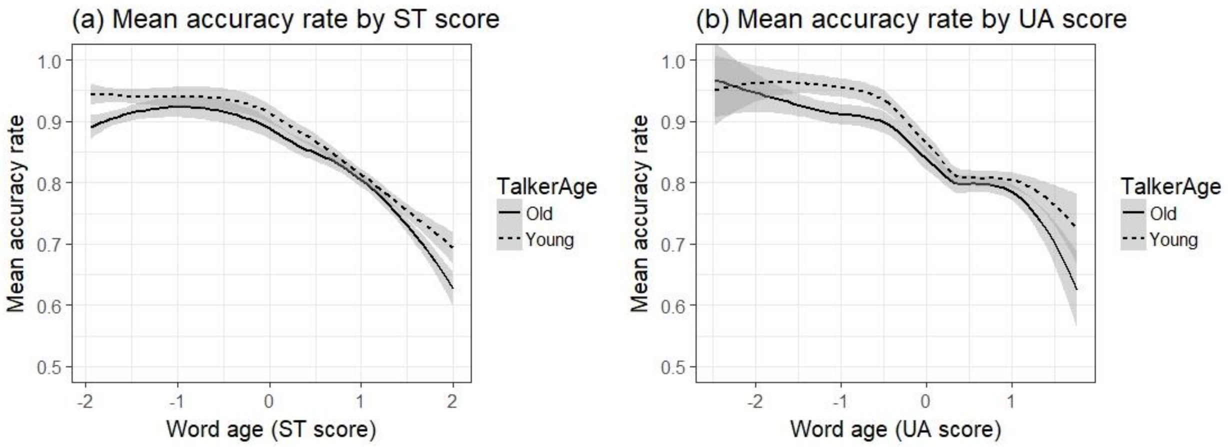


Figure 2.3. Mean accuracy rates by word age and talker age: Word age is represented by the ST score in (a) and by the UA score in (b). Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

Evident in both graphs are main effects of talker age and word age; participants tended to recognize the words less accurately (1) when words were produced by older talkers than younger talkers, and (2) as word age increased. These effects are due to the young ages of the participants; they are unfamiliar with the phonetic realizations of older talkers and old-associated words. Importantly for the hypothesis, however, words were not identified more accurately when talker age matched word age. Although accuracy rates predicted by ST score show the predicted pattern for words with ST score between -2 and 1, highly old-associated words did not benefit from older voices.

As outlined above, I ran two frequentist models fit to accuracy with either ST or UA score as a predictor of word age in each model, in order to test the effects of stereotypes and usage separately. Then, the effects of the two predictors will be compared using model comparisons in Section 2.1.2.3.

To test the influence of stereotype-based word age on accuracy statistically, a binomial mixed effects model was fit to the binary accuracy data. First, a series of frequentist models tested various fixed effects, including trial order, test block order, test location, participant sex, presentation order of talker age, item list, talker gender, and word duration. Each of these factors was added along with the three test variables, (i.e., talker age, word age (ST score), and their interaction), and all two-way and three-way interactions were also tested. And then, only ones that reached significance and improved the model's fit were included in the final model. Model fit was evaluated using analysis of variance (ANOVA). All two-way and three-way interactions between each factor and the test variables were tested. Included in the final model as fixed effects (for both the frequentist and Bayesian models) were talker age (older or younger, deviation coded), ST score (continuous), an interaction between them, and talker gender (deviation coded).

To define the random effects structure of the frequentist model, I began maximally, trimming variables that either failed to converge or failed to improve the model's fit. As a result, by-item intercepts, by-participant intercepts, and by-participant slopes for talker age were included in the final model.

Table 2.2. Summary of frequentist model fit to accuracy with ST score as word age: High ST score indicates that the word is stereotypically associated with older talkers. Negative coefficients indicate low accuracy.

Model: glmer (Accuracy ~ TalkerAge*ST + TalkerGender + (1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	2.476	0.122	20.333	<.001
Talker age=young	0.360	0.094	3.836	<.001
ST score	-0.508	0.069	-7.331	<.001
Talker gender=male	-0.153	0.055	-2.763	.006
Talker age=young : ST score	-0.141	0.044	-3.192	.001

As shown by the positive estimated coefficient for talker age in Table 2.2, words spoken by younger talkers were significantly more frequently recognized correctly than words spoken by older talkers ($p<.001$). There was also a significant main effect of ST score; as ST score increased, participants were less likely to respond correctly ($p<.001$). Regarding the effect of talker gender, words spoken by males were less frequently recognized correctly than by females ($p<.01$)[17]. Importantly, there appears a significant interaction between talker age and word age; when words were heard in younger talkers' voice, in comparison with older talkers' voice, participants were less likely to respond correctly for words with higher ST scores ($p<.01$).

For comparison, the effect of usage-based word age was examined in a separate binomial mixed effects model, by replacing the predictor of word age from ST score to UA score. All other predictors tested in this model were identical with the model predicted by ST score.

Table 2.3. Summary of frequentist model fit to accuracy with UA score as word age: A word with a high UA score is one that was reported to be spoken more frequently by older speakers.

Model: glmer (Accuracy ~ TalkerAge*UA + TalkerGender + (1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	2.533	0.125	20.333	<.001
Talker age=young	0.362	0.095	3.817	<.001
UA score	-0.654	0.107	-6.135	<.001
Talker gender=male	-0.153	0.055	-2.771	.006
Talker age=young : UA score	-0.185	0.070	-2.645	.008

---

[17] This effect is also found in other models reported in this chapter; words produced by male talkers were identified less accurately and more slowly. As will be shown in Section 2.3, durations of the acoustic stimuli were shorter for male talkers than female talkers. The shorter durations for male talkers seem to be induced by the older male talker's tendency to produce words in a reduced form, which resulted in decreased recognition facility for male talkers in general. Thus, I interpret that the gender effect on accuracy and RT is not generalizable but arose from idiosyncrasies specific to the talkers I used.

As shown in Table 2.3, there were significant main effects of talker age, talker gender, and UA score in the same directions as in the models that included ST score (cf., Table 2.2). That is, words were recognized more accurately when spoken by younger talkers (p<.001) and less accurately when spoken by male talkers (p<.01), and words with higher UA scores were less likely recognized correctly (p<.001). The interaction between talker age and word age is also significant in this model, indicating that recognition accuracy was decreased when words with higher UA scores were spoken by younger talkers (p<.01)[18].

So far, this section has demonstrated that either ST or UA score interacts with talker age significantly, respectively in a separate model fit to the accuracy data. The effects of the two word age predictors on both accuracy and reaction times will be compared using a model comparison method in 2.1.2.3.

2.1.2.2. Reaction times. Reaction times (RTs) were measured from the beginning of a word to the response (button press). Only reaction times for correct responses to real words (16,033 tokens) were included in the analysis. Data points below 2/3 of the target word duration and over 5,000ms were removed first (N=7), and then responses that fall out of three standard deviations from the mean by participant were removed (N=178). This procedure left 15,848 tokens for the analysis, excluding 185 outliers (1.15% of the total). Mean RTs of the critical trials (excluding filler trials, N=11,550) are plotted by talker age and the two types of word age in Figure 2.4.

---

[18] When the two models fit to accuracy (Table 2.2 and 2.3) were refitted with the maximal random effects structure, both models failed to converge and were nearly unidentifiable with a large eigen value.

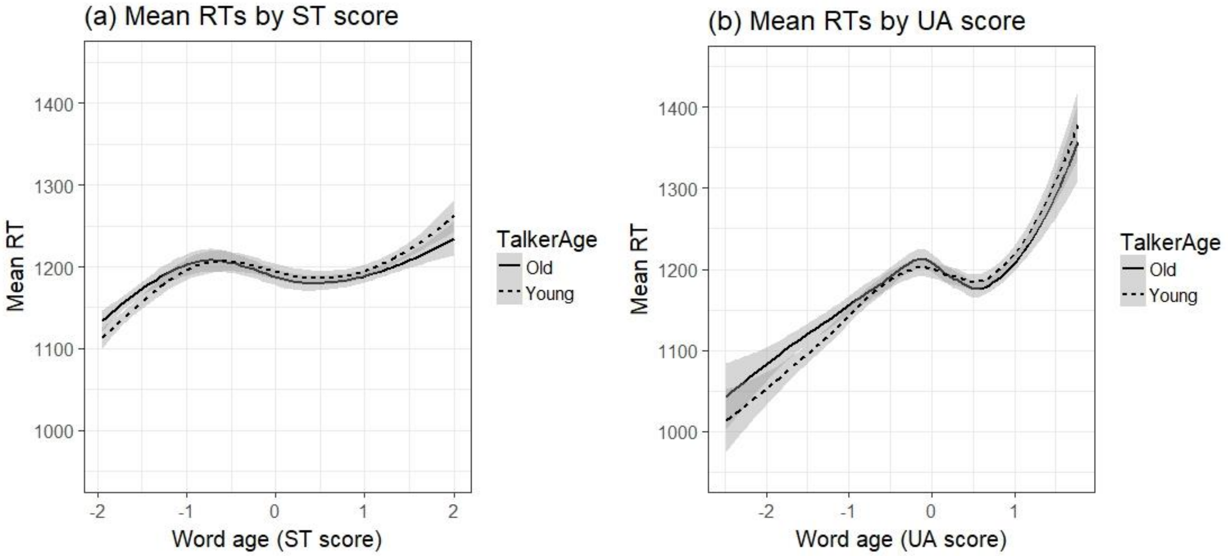


Figure 2.4. Mean RTs by word age and talker age: Word age is represented by the ST score in (a) and by the UA score in (b). Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

Both graphs indicate a main effect of word age; participants took longer to recognize a target as a real word, as word age increased. Also evident is a cross-over interaction between word age and talker age. That is, words with lower word ages were recognized faster when produced by younger talkers, while recognition of words with higher word ages was facilitated when produced by older talkers. This appears to be the case when either the ST or UA score is used to represent word age.

It is also notable that responses for both older and younger talkers were substantially slowed down for particular items that are slightly associated with younger talkers, indicated by a hump around the ST word age region between -1 and -0.5, and around the UA word age region between -0.4 and 0. This seems to be due to an effect of word duration, and – since the experiments use the same critical items – it is observed for all of the lexical decision experiments reported in this dissertation. The mean duration of all critical items is 765ms. However, durations are longer for the 10 items with ST score from -1 to -0.5 (mean=972ms, min=825, max=1,175), all of which are accidentally tri- or quadri-syllabic words. Durations of the 23 items with UA score from -0.4 and 0 are also substantially longer than the overall mean (mean=881ms, min=439, max=1,278).

To test whether the trends described above are statistically significant, I present two frequentist models fit to RTs (one with ST score and another with UA score). First, a linear mixed effects model was fit to raw RTs with ST score as the word age variable. Through the same process reported in Section 2.1.2.1, fixed effects in the final model were main effects and interaction of talker age (older or younger, deviation coded) and ST score (treated as continuous), word duration (as a control variable), talker gender (deviation coded), and trial order (continuous). In this model, the maximal random effects structure was fit without failure to converge using lme4, including (1) random intercepts for by-participant and by-item variance, (2) by-participant random slopes for talker age, ST score, and their interaction, and (3) by-item random slopes for talker age and an interaction between talker age and ST score.

Table 2.4. Summary of mixed effects model fit to RT with ST score as word age: A high ST score indicates that the word is highly associated with older talkers. Positive coefficients indicate increased RTs.

Model: lmer (RT ~ TalkerAge*ST + Duration + TalkerGender + TrialOrder +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	685.913	20.426	33.580	<.001
Talker age=young	-13.553	8.045	-1.685	.092
ST score	24.726	4.465	5.538	<.001
Duration	0.654	0.021	31.214	<.001
Talker gender=male	8.250	3.407	2.421	.015
Trial order	-0.030	0.014	-2.114	.035
Talker age=young : ST score	9.238	3.627	2.547	.011

As shown in Table 2.4, the younger talker trials tended to be responded to faster than the older talker trials, but the difference was not significant (*p*=.092). There is a significant main effect of ST score; as ST score increased, participants recognized the words more slowly (*p*<.001). A significant main effect of word duration indicates that RTs increased for longer words (*p*<.001), and therefore duration was included as a control variable. As indicated by a significant main effect of talker gender, responses were slower for words spoken in male voices than in female voices (*p*<.05). RTs are also influenced by trial order; words are recognized faster, as the experiment session progressed (*p*<.05). Additionally, there is a significant interaction between talker age and stereotypical word age (ST score); as evident in Figure 2.4 (a), when words were heard in a younger talker's voice, in comparison with an older talker's voice, participants were slower to recognize words with higher ST scores (*p*<.05).

UA score was also tested as a fixed effect (instead of the ST score) in a separate model fit to raw RTs of the same dataset, by replacing the predictor of word age with UA score. All other predictors tested in this model were identical with the RT model predicted by ST score.

Table 2.5. Summary of mixed effects model fit to RT, with UA score as word age: A high UA score indicates that the word was reported as being more often used by older speakers than younger speakers.

Model: lmer (RT ~ TalkerAge*UA + Duration + TalkerGender + TrialOrder +
(1+TalkerAge*UA | Participant) + (1+TalkerAge:UA+TalkerAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	690.879	20.329	33.984	<.001
Talker age=young	-14.976	8.012	-1.869	.062
UA score	32.334	5.901	5.480	<.001
Duration	0.645	0.021	30.790	<.001
Talker gender=male	7.739	3.409	2.270	.023
Trial order	-0.030	0.014	-2.104	.035
Talker age=young : UA score	14.058	4.999	2.812	.005

As shown in Table 2.5, the effects found in the UA score model are comparable with those in the ST score model. Participants tended to recognize words faster when the words were heard in younger talkers' voices than in older talkers' voices, but the effect did not reach significance (*p*=.062). Participants took longer to recognize words with higher UA scores (*p*<.001), or with longer durations (*p*<.001). Words spoken in male voices were responded to more slowly than words spoken in female voices (*p*<.05). RTs were shorter for trials that appeared later in the experiment session (*p*<.05). As in the model with stereotype age, the interaction between talker age and word age is also significant (*p*<.01) when UA score is included in the model as a predictor of word age, indicating that word age based on self-reported frequency is also an efficient predictor of links between words and age groups.

2.1.2.3. Comparison between ST and UA. The above analyses demonstrate that recognition accuracy and speed for age-associated words are improved when the word is spoken by an age-congruent talker, and that this effect is predicted by either stereotype- or usage-based word age. Recall that the two types of word age predictors are statistically correlated, but also have substantial variation for some lexical items (see Figure 2.1). To examine whether stereotypes wield an influence on lexical recognition above and beyond the effect of frequency of use, this section tests whether ST score accounts for a greater amount of variance than UA score does through a set of model comparisons.

First, a preliminary linear regression model was fit to predict the ST scores of critical items with their UA scores. The residuals from this model, or residualized ST, constitutes a variable that represents the variance of ST scores independent from the variance explained by UA scores. Then, residualized ST is added as a predictor into the aforementioned UA score model (reported in Table 2.5), to test the degree to which the sum of residuals in the UA score model is predicted by residualized ST when it is added as a predictor variable for word age in addition to UA score[19]. However, as the model including residualized ST failed to converge with the maximal random effects structure, the model was refitted excluding the by-item random slope for talker age. In this model, the main effect of residualized ST remained significant (β=21.451, s.e.=8.260, *p*=.009), indicating that an effect of ST score isolated from that of UA score still accounts for a substantial amount of residual variance from the UA model. Further, an ANOVA comparison between this model and a UA model refitted with an equivalent non-maximal random effect structure reveals that the addition of residualized ST improves the fit of the model with UA score only ($\chi^2(1) = 6.761$, *p*=.009).

The second model comparison was performed for UA score residualized on ST to dissociate the effect of ST from UA. The results from a model that included residualized UA and raw ST, along with non-maximal random effects structure excluding the by-item random slope for an interaction between word age and talker age, revealed that residualized UA neither remained significant (β=7.787, s.e.=12.496, *p*=.533), nor improved the fit of the model that included ST score (Table 2.4) ($\chi^2(1)=0.53$, *p*=1). The two comparisons, therefore, demonstrate that ST score has an additional effect that is isolated from an effect of UA score, but not vice versa[20].

---

[19] This method is used in psycholinguistic analyses to evaluate the effects of correlated predictors (e.g., Kuperman, Bertram, & Baayen, 2010; Cohen-Goldberg, 2012).

[20] Similar results were found in non-converging models with the maximal random effect structure. In a model that included residualized ST and raw UA, residualized ST significantly influenced RTs (β=20.164, s.e.=8.284, *p*=.015). In a model that included residualized UA and raw ST, a main effect of residualized UA was not significant (β=10.150, s.e.=12.130, *p*=.403).

Similar results were also found in model comparisons fit to accuracy. Residualized ST remained significant (β=-0.498, s.e.= 0.133, *p*<.001) and improved the fit of the UA model ($\chi^2(1) = 13.431$, *p*<.001), while residualized UA did not (β=0.001, s.e.= 0.198, *p*=.997), ($\chi^2(1) = 0$, *p*=1). These models only included random intercepts for by-participant and by-item variation due to convergence errors.

2.1.2.4. Analysis with Bayesian fitting. The results presented above demonstrate that both accuracy and speed of word recognition are improved when the talker's age matches word age derived from either stereotypes or usage, and that stereotypes wield a greater influence than usage. However, recall that the random effects structure in the frequentist model fit to accuracy (Table 2.2) is not maximal, omitting (1) by-item slopes for main effects and interaction of talker age and word age, and (2) by-participant slopes for word age and an interaction between word age and talker age. This raises a concern because omitting random slopes can inflate the false-positive rate (Barr et al., 2013).

To further investigate the accuracy results in conservative statistical inference, I fit a Bayesian hierarchical model, which enables me to run the model including the maximal random effects structure without failure to converge. Because the RT model with the maximal structure reveals a significant effect (Table 2.4 and 2.5), and because ST score appears to be a more reliable predictor for participant behavior (both accuracy and RT), I only present a model fit to accuracy including ST score as a word age variable in this section.

In Bayesian fitting, we inform the model of *priors*, i.e., predicted distribution of the parameters, and the model presents *posterior* values for the parameters by computing a weighted mean of the prior and the likelihood from the given data. Then, the effects of the predictors are evaluated by the posterior values and their probabilities in the sampled posterior distribution

(Nicenboim & Vasishth, 2016, see also Hofmeister & Vasishth, 2014; Husain, Vasishth, & Srinivasan, 2014; Frank et al., 2016, for application of this method for psycholinguistic experiments). In addition to conservatism, one important advantage of this method is that evaluation of the effect based on the posterior probability of a parameter (e.g., the probability of the coefficient being less/greater than 0) is more direct evidence for the hypothesis at hand than the traditional paradigm that only determines significance of the effect based on *p*-value, a conditional probability that the null hypothesis is true (Husain et al., 2014).

A Bayesian model was fit to the accuracy data in Experiment 1 with the same fixed effects in the frequentist model above (Table 2.2), which are talker age, word age (ST score), an interaction between them, and talker gender. The maximal random effects in the model included (1) by-item intercepts, (2) by-participant intercepts, (3) by-participant slopes for talker age, word age, and their interaction, and (4) by-item slopes for talker age and the interaction between word age and talker age.

As for priors setting, when the sample size is large enough, it is normal to insert weakly informative priors, because precisely specified priors will have an excessive influence on determining the posterior distribution (Nicenboim & Vasishth, 2016). The priors were set as normal(0, 10) (i.e., a normal distribution with a mean of 0 and a standard deviation of 10) for the intercept, and normal(0, 1) for the fixed effects. The prior for the regularization of the covariance in random effects was set to 2. These priors were justified to be weakly informative by the fact that the posterior values did not change substantially when varying weakly informative prior specifications were applied.

The data was simulated with four chains and 2,000 iterations, each out of which 1,000 iterations were warm-up samples. The model was checked for proper chain mixing. The Gelman-

Rubin convergence statistics (r-hat values) of all predictors were close to or equal to 1, indicating that the models' chains converged (Gelman & Rubin, 1992).

Table 2.6. Summary of Bayesian model fit to accuracy: Mean posterior estimates and standard deviation are provided for each factor, as well as the 2.5% and 97.5% boundaries of the credible interval and P(β<0), i.e., the probability of the coefficient being lower than 0.

Model: stan_glmer (Accuracy ~ TalkerAge*ST + TalkerGender +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

Factor	Mean	SD	Credible interval		P(β<0)
			2.5%	97.5%	
(Intercept)	2.455	0.131	2.208	2.724	0
Talker age=young	0.390	0.127	0.149	0.636	.001
ST score	-0.403	0.090	-0.585	-0.234	1
Talker gender=male	-0.158	0.056	-0.271	-0.045	.997
Talker age=young : ST score	-0.142	0.094	-0.326	0.041	.930

As summarized in Table 2.6, the main effects of talker age, ST score, and talker gender in the Bayesian fitting are robust and comparable to the frequentist model; response accuracy increased when listening to younger talkers than older talkers, and decreased for words with higher ST score and for male talkers. The interaction between talker age and word age also trended in the same direction but was subtle when the maximal random effects were accounted for; with 93% probability, words produced by younger talkers were recognized less accurately as ST score increased.

2.1.2.5. Summary of results. In Experiment 1, I have shown that word recognition facility increased when the word is spoken by a talker whose age is congruent with the age group that is either reported to frequently produce the given word (measured by UA score), or is stereotypically associated with the word (measured by ST score). Although the effect on accuracy is significant only in a frequentist model with a non-maximal random effects structure (for either ST or UA scores), a robust effect is found for both predictors when recognition speed is tested. Additionally, the effect is better predicted by stereotypes than usage. As will be discussed in Section 2.4, these results highlight the role of socio-indexical information in the signal during lexical access.

2.2. Experiment 2: Rapid influence of phonetic detail on lexical access

Results from Experiment 1 are consistent with Walker and Hay's (2011) finding that word recognition is facilitated by congruence between word age and talker age. In both experiments, however, stimuli were blocked by talker, meaning that participants heard each talker's voice continuously across all trials in a given block. Thus, the voice heard in previous trials within a block provided listeners with ample time and phonetic information, with which they could form expectations about the talkers or even about which types of words the talker most likely uses. In this sense, even though words were spoken in isolation, the voice itself could have provided more or less contextual information influencing the probability of particular lexical items to be activated.

Experiment 2 explores the possibility that the effect in Experiment 1 arises from listener expectations held prior to word onset. To test this, Experiment 2 replicates Experiment 1 but uses a mixed presentation of talkers instead of blocking by talker. As the two experiments differ only in whether stimuli are presented in blocked or mixed design, a comparison of effect size can be performed to examine whether preexisting expectations influence lexical access above and beyond any more immediate effects from phonetic detail in the target word. If the same results are found and the effect size is comparable even when the phonetic cues to talker age are only present within the target word, it would suggest that the processing advantage driven by age-congruence arises from a bottom-up process, in which pre-existing expectations are not required but phonetic detail in the signal by itself is sufficient to guide access to socially-indexed lexical items.

2.2.1. Method

The method used in Experiment 2 is identical to Experiment 1, except that the stimuli were not blocked by talker. Lexical items were counterbalanced by four talkers, and were distributed across four lists in a Latin Square design. Each participant was randomly assigned to one of the four item lists, and items appeared in a random order for each participant. Each participant was able to choose to take a short break after every 1/4 of the experiment session (192 trials).

The data were collected from 32 participants (22 females and 10 males, age: 18-25) in two different locations: 17 participants took part at the University of Hawaiʻi at Mānoa and 15 at Chung-Ang University in Seoul, Korea. All 32 participants were registered students at colleges in the Seoul Metropolitan area, and all listed the Seoul Dialect as their first or second most frequently used dialect. The Hawaiʻi participants had been in Hawaiʻi for less than four weeks at the time of participation. No difference in behavior between the two locations was either hypothesized or observed. Across the four experiments reported in this dissertation, no participants took part in more than one experiments.

2.2.2. Results

In this section, results from Experiment 2 are reported and compared with Experiment 1 to examine whether any difference in participant behavior is observed when stimuli are presented without talker block. For simplicity, ST score (rather than UA score) is used to represent word age, henceforth. This is because ST score better predicts the participant behavior in Experiment 1, and this trend was also confirmed through preliminary analyses for all experiment data reported in this dissertation.

Just as in Experiment 1, I begin this section by presenting a frequentist model fit to accuracy (Section 2.2.2.1) and RTs (2.2.2.2). Next, an effect size comparison is conducted by fitting a frequentist model that includes presentation mode (mixed vs blocked) as a fixed effect (2.2.2.3). Finally, two Bayesian models (for accuracy and RT, respectively) are presented to support results from the frequentist analysis (2.2.2.4).

2.2.2.1. Accuracy. Among the 24,576 tokens gathered from 32 participants, there were 20,761 correct responses. The accuracy rate is 85% for real words and 84% for non-words. Mean accuracy rates of the real-word targets from the current experiment (9,216 tokens excluding fillers) are plotted by word age (ST score) and talker age in Figure 2.5 (b). For comparison, the accuracy data from Experiment 1 are repeated in Figure 2.5 (a) (13,824 tokens).

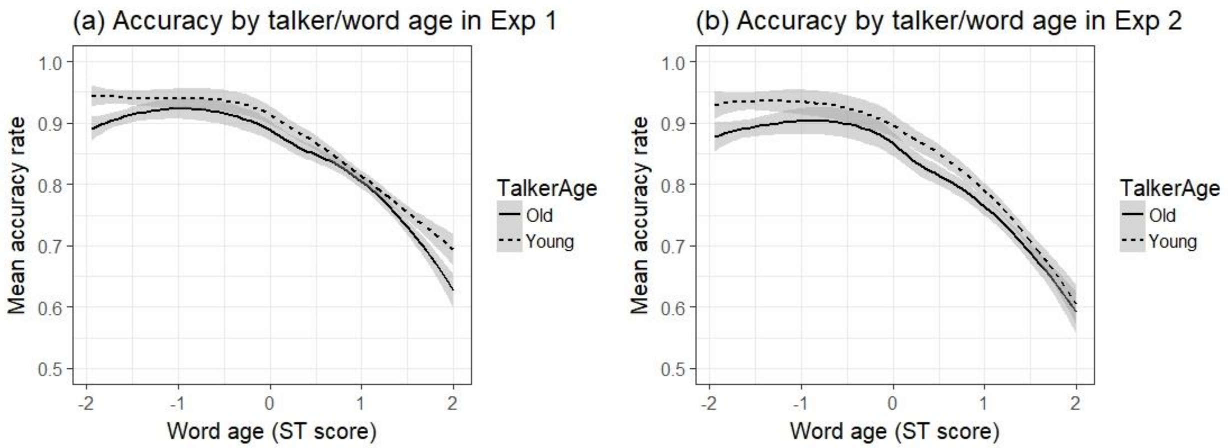


Figure 2.5. Mean accuracy by word age and talker age: (a) data from Experiment 1 where stimuli are blocked by talker, (b) data from Experiment 2 where stimuli are presented randomly. Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

The main effects of talker age and word age are found in both graphs: the young listeners more accurately recognized (1) words produced by younger talkers than by older talkers, and (2) words with lower word age. In the data from Experiment 2, the interaction between talker age and word age is shown; the advantage for younger voices, indicated by the difference in mean accuracy across talker age, is greater for words with lower word age, suggesting that recognition of young words is boosted by younger voices beyond the main effect of talker age. However, a cross-over interaction is not found for accuracy in either experiment; it is not the case that accuracy for old words also benefit from the age-congruence effect.

To explore the results further, a frequentist binomial mixed effects model was fit to accuracy. Through the process reported in 2.1.2.1, talker age (deviation coded), word age (continuous), and their interaction were included as fixed effects in the final model, and a non-maximal random effects structure was fit in the model including by-item intercepts, by-participant intercepts, and by-participant slopes for talker age.

Table 2.7. Summary of frequentist model fit to accuracy in Experiment 2

Model: glmer (Accuracy ~ TalkerAge*ST + (1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	2.284	0.143	15.927	<.001
Talker age=young	0.392	0.075	5.234	<.001
Word age (ST score)	-0.579	0.074	-7.843	<.001
Talker age=young : Word age	-0.182	0.052	-3.491	<.001

As shown in Table 2.7, the main effects of talker age and word age indicates that words were identified more accurately when produced by younger talkers (*p*<.001), and less accurately as word age increased (*p*<.001). In addition, a significant interaction is found between the two factors; when words were produced by younger talkers, compared to older talkers, words were less correctly identified for words with higher word age (*p*<.001), confirming the pattern of mean accuracy rates shown in Figure 2.5 (b)[21].

2.2.2.2. Reaction times. Reaction times were measured and went through the outlier treatment as reported in 2.1.2.2. As a result, 158 tokens were removed from the total of 10,401 tokens (<2%). Also excluding data from filler trials, RT data for trials with critical items (7,432 tokens, M=1,178, SD=249.54) are plotted in Figure 2.6 (b), along with the data from Experiment 1 (11,550 tokens, M=1,182, SD=229.66) in Figure 2.6 (a).

---

[21] The same trend was found in a separate frequentist model with the maximal random effects structure, which failed to converge. The interaction between talker age and word age in that model showed significantly lower accuracy when words with higher word age were produced by younger talkers than by older talkers (β=-.333, s.e.=.107, *p*=.002).

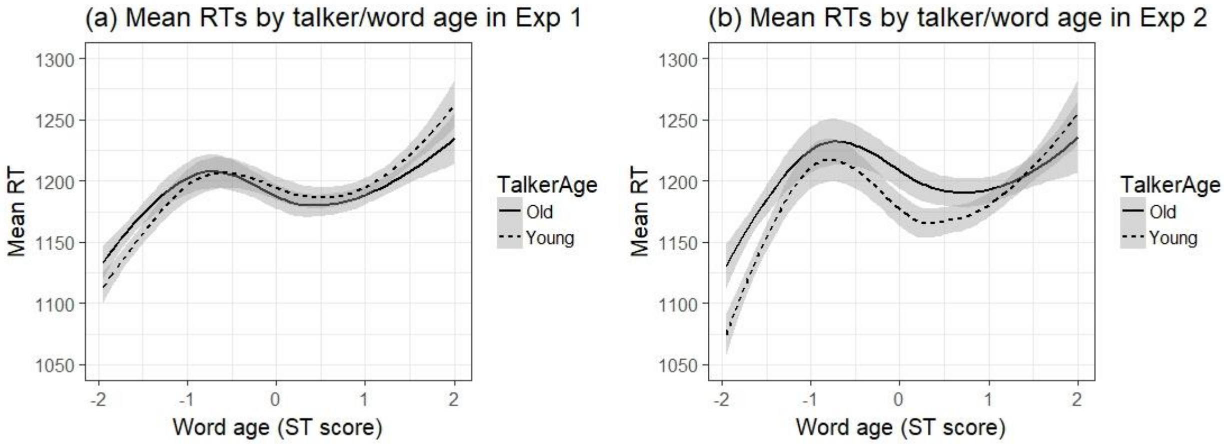


Figure 2.6. Mean RTs by word age and talker age: (a) data from Experiment 1 where stimuli are blocked by talker, (b) data from Experiment 2 where stimuli are presented in random order. Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

There are some notable trends found in both graphs. First, responses to words with word ages between -1 and -0.5 were slowed down due to them having longer word durations. Second, responses were faster for words with lower word age. As for the effect of talker age, a cross-over interaction is shown in both datasets, indicating that recognition of either old- or young-associated words was facilitated by age-congruent voices, although the point where the two lines cross is not consistent across experiments. The difference in skewness seems to suggest that the two different groups of participants benefitted from age-congruence among slightly different word ages. Also, listeners seem to have benefitted from the younger voices to a greater degree in general in Experiment 2 than in Experiment 1 – indicated by greater difference in RTs across talker age in 1(b). However, further work is needed to tell whether these effects are induced by the difference in design (i.e., mixed vs blocked presentation of talkers).

A frequentist mixed effects model was fit to the RT data from the current experiment. Using the same procedure described in 2.1.2.1, the final model included talker age (deviation coded), word age, an interaction between them, word duration, talker gender (male or female, deviation coded), and trial order (continuous) as fixed effects. In addition, the maximal random effects structure was included without failure to converge in this model.

Table 2.8. Summary of frequentist model fit to RT in Experiment 2

Model: lmer (RT ~ TalkerAge*ST + Duration + TalkerGender + TrialOrder +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	702.058	26.487	26.506	<.001
Word duration	0.640	0.025	25.843	<.001
Talker age=young	-30.952	6.382	-4.850	<.001
Word age (ST score)	24.462	4.899	4.993	<.001
Talker gender=male	12.271	4.464	2.749	.006
Trial order	-0.040	0.010	-4.130	<.001
Talker age=young : Word age	7.835	4.345	1.803	.071

Shown in Table 2.8 are main effects of word duration, talker age, word age, and talker gender, all of which are trending in the same directions as in Experiment 1; listeners took a longer time to recognize words that were (1) longer in duration (*p*<.001), (2) associated with older people (*p*<.001), and (3) produced by male talkers rather than female talkers (*p*<.01). Likewise, slower recognition occurred for words produced by older talkers compared to those produced by younger talkers (*p*<.001). A main effect of trial order is also significant, indicating that participants responded more and more quickly as the experiment session proceeded (*p*<.001). In contrast with the results from Experiment 1, the interaction between talker age and word age is not significant in Experiment 2 (*p*=.071); the processing advantage from age-congruence between the talker and the word appears to be reduced when stimuli are not blocked by talker, although it is trending in the expected direction.

2.2.2.3. Comparison between blocked and mixed presentations. Recall that the design of the two experiments reported in Chapter II differs only in how talkers from different age groups are presented across stimuli. This allows me to compare the effect size across the experiments and test to what degree the age-congruence effect found in Experiment 1 is induced by pre-existing expectation about the talker. By fitting a mixed effects model to a combined dataset of the two experiments that includes presentation mode (mixed vs. blocked) as a predictor, this section tests whether there is any difference in participant behavior; particularly, if prior expectation has an additional effect above the effect driven by the signal, then recognition may be facilitated by the age-congruence effect to a greater degree when the stimuli are blocked by talker.

A linear mixed effects model was fit to RTs from a combined dataset of both experiments. The fixed effects in the model included a 3-way interaction of talker age (deviation coded), word age, and presentation mode (mixed vs. blocked, deviation coded). Other predictors were identical to the RT model.

Table 2.9. Summary of RT model fit to data from two experiments: A 3-way interaction is included to test the influence of talker presentation mode on the age-congruence effect.

Model: lmer (RT ~ TalkerAge*ST*Mode + Duration + TalkerGender + TrialOrder + (1+TalkerAge*ST|Participant) + (1+TalkerAge:ST+TalkerAge|Item))

	Estimate	Std. Error	t value	p value
(Intercept)	667.491	19.597	34.060	<.001
Talker age=young	-14.083	6.881	-2.047	.041
Word age (ST score)	25.272	4.708	5.368	<.001
Mode=mixed	1.737	17.196	0.101	.920
Word duration	0.681	0.019	36.145	<.001
Talker gender=male	11.584	2.753	4.209	<.001
Trial order	-0.034	0.007	-4.611	<.001
Talker age=young : Word age	9.345	3.637	2.570	.010
Talker age=young : Mode=mixed	-17.538	9.396	-1.867	.062
Word age : Mode=mixed	0.635	2.800	0.227	.821
Talker age=young : Word age : Mode=mixed	-1.704	4.058	-0.420	.675

As shown in Table 2.9, the effect of presentation mode is not significant either as a main effect (*p*=.920), or in interaction with other factors. The mixed design neither influenced listeners' sensitivity to word age significantly (*p*=.821), nor did it influence the effects of congruence/incongruence between talker age and word age (*p*=.675). A non-significant interaction between talker age and presentation mode indicates a slight tendency for participants in the mixed presentation mode to respond faster to voices of younger talkers (*p*=.062). It is also notable that congruence between talker age and word age also significantly facilitated lexical access in the combined dataset (*p*<.05).

Additionally, the effect of presentation mode was not significant either when accuracy data between the two experiments was compared using the same method. Overall accuracy rate was decreased in the mixed mode than in the blocked mode but this effect did not reach significance (*p*=.060). Presentation mode did not interact with talker age (*p*=.705), with word age (*p*=.174), or with congruency between talker age and word age (*p*=588). Along with these effects, an interaction between talker age and word age was found, indicating that listeners were more accurate when the two factors were congruent (*p*<.001).

In sum, the effect size comparison shows that the age-congruence effect is not significantly affected, in terms of both RT and accuracy, by whether talkers are presented within a block or randomly, suggesting that prior expectation plays little role in the observed effect. In addition, since a significant interaction between talker age and word age is still found in the comparison model, the non-significant effect on RTs in Experiment 2 does not seem to be due to the difference in design, but possibly due to a smaller sample size (48 participants in Experiment 1, and 32 in Experiment 2).

2.2.2.4. Analysis with Bayesian fitting. The frequentist analysis above demonstrates trends for improved recognition accuracy and speed when talker age matches word age. However, both accuracy and RT results are worth further investigation because (1) the accuracy model is not maximal, and (2) the interaction does not reach significance in the RT model. Therefore, I reexamine the data in this section with Bayesian models which (1) enable to fit the maximal model without failure to converge, and (2) operate under different statistical inference; specifically, Bayesian models make use of priors and given data to obtain the posterior probability of the hypothesis being true (see Section 2.1.2.4).

To examine the accuracy data with random slopes for all test variables, a Bayesian hierarchical model was fit to the same accuracy dataset presented in 2.2.2.1. Fixed effects in the model were identical to the frequentist model (Table 2.7), including talker age, word age, and their interaction. The random effects structure was maximal and identical to the Bayesian accuracy model reported in Experiment 1, including (1) by-item intercepts, (2) by-participant intercepts, (3) by-participant slopes for talker age, word age, and their interaction, and (4) by-item slopes for talker age and an interaction between word age and talker age. The settings for the priors and simulation were also identical to the Bayesian accuracy model in Experiment 1. The priors and chain mixing were checked through the same process described in 2.1.2.4.

Table 2.10. Summary of Bayesian model fit to accuracy in Experiment 2: Mean posterior estimates and standard deviation are provided for each factor, as well as the 2.5% and 97.5% boundaries of the credible interval and P(β<0), i.e., the probability of the coefficient being lower than 0.

Model: stan_glmer (Accuracy ~ TalkerAge*ST +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

Factor	Mean	SD	Credible interval		P(β<0)
			2.5%	97.5%	
(Intercept)	2.08	0.15	1.79	2.38	0
Talker age=young	0.58	0.13	0.34	0.82	0
Word age (ST score)	-0.37	0.09	-0.56	-0.19	1
Talker age=young : Word age	-0.32	0.10	-0.52	-0.12	1

The posterior values obtained from a Bayesian fitting indicate similar effects found above, both in Experiment 1 and in the frequentist model of Experiment 2. Shown in Table 2.10 are (1) higher accuracy for younger talkers, (2) lower accuracy as word age increases, and (3) lower accuracy for words with higher word age produced by younger talkers with 100% probability.

Next, the RT data were also tested with Bayesian fitting to further examine the results provided in Section 2.2.2.2. Fixed effects and random effects included in the model were identical to the frequentist RT model (Table 2.8). For the Bayesian RT model, I used the R package *brms* (Bürkner, in press) because it was confirmed to sample the given data more efficiently than rstanarm during sampling check using the function *launch_shinystan*. Priors for the fixed effects and their standard deviation were both set to normal(0, 10). The prior for the correlation of the group level random effects (i.e., lkj) was set to 2. Other settings for the sampling were equal to the accuracy model. Posterior values were not substantially affected by varying weakly informative priors, and r-hat values of all predictors were close to or equal to 1.

Table 2.11. Summary of Bayesian model fit to RT in Experiment 2: This model was fit using the brms package.

Model: brm (RT ~ TalkerAge*ST + Duration + TalkerGender + TrialOrder +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

Factor	Mean	SD	Credible interval		P(β<0)
			2.5%	97.5%	
(Intercept)	715.84	22.11	672.90	759.68	0
Word duration	0.62	0.02	0.58	0.67	0
Talker age=young	-22.19	4.73	-31.34	-12.92	1
Word age (ST score)	16.81	4.41	8.05	25.25	0
Talker gender=male	10.03	4.08	2.10	17.93	.005
Trial order	-0.04	0.01	-0.06	-0.02	1
Talker age=young : Word age	8.28	3.87	0.78	15.83	.014

In Table 2.11, along with similar patterns of all fixed effects with the frequentist model, the posterior for the interaction between talker age and word age reveals with 98.6% probability that words produced by younger talkers were recognized more slowly as word age increased, providing supporting evidence that incongruence between the two predictors delays recognition speed. Thus, while the interaction is merely approaching significance in the frequentist model, the Bayesian analysis suggests that the interaction is not due to chance.

2.3. What acoustic properties contribute to the effect?

Results from the two experiments reported in Sections 2.1 and 2.2 demonstrate that word processing is facilitated by congruence between talker age and word age. While this finding highlights the role of socially-indexed phonetic realizations in lexical access, a question that ensues is what specific features in older and younger speakers' speech induced the effect. As Walker and Hay (2011) note, it could have been caused by a combination of different acoustic properties that cue the ages of speakers. However, as reviewed in Section 1.3, variability in phonetic realizations of words is predictable to the extent that the distribution of word usage is constrained by social factors.

For example, high-frequency words tend to be produced with lenitive forms and thus with shorter durations (Bybee, 2001; 2002). Since such processes are also intertwined with socially-conditioned distributional factors in word usage (Hay & Foulkes, 2016), it is possible that realizations of phonetic details contributing to speech rate (e.g., lenition versus fortition) are influenced by individual talkers' experiences with the words. If so, an effect of age congruence on phonetic reductions may be observed in the stimuli used in my experiments, even though I intended control prosodic details at the time of recording (see Section 2.1.1.4). Specifically, the lexical stimuli may have been produced with shorter durations when produced by talkers belonging to the age group who produce the word frequently (e.g., young words produced by younger talkers) than when the words were produced by talkers from an incongruent age group.

It is also conceivable that such age-related variability in production details would be present in multiple acoustic domains (e.g., variation in pitch, intensity, or realization of phonetic variables). However, as a post-hoc analysis for speech tokens obtained from a limited number of individuals, I focus on the age-related variability in the temporal domain, testing the degree to which word duration was influenced by the congruence between talker age and word age (Section 2.3.1). Then, I test whether the effects of the effects found in Experiments 1 and 2 are still predicted by perceptual congruence in age-indexed phonetic realizations, even when the age-related temporal variability is filtered out (2.3.2).

2.3.1. Talker-specific variability in word duration

Recall that the duration of the auditory stimuli was a robust predictor in all RT models reported in Sections 2.1 and 2.2, showing that participants took longer time to respond to words with longer duration. While those analyses treated the durational variability as if it is a word-specific factor inherently conditioned by the length of the word form (i.e. number of syllables/segments), the purpose in the current analysis is to investigate the degree to which word duration is predicted by an interaction between talker age and word age, along with other factors pertaining to the speech tokens obtained in this dissertation. Figure 2.7 presents distributions of durations of the critical items (i.e., real word items that had been assigned word age) across talkers.

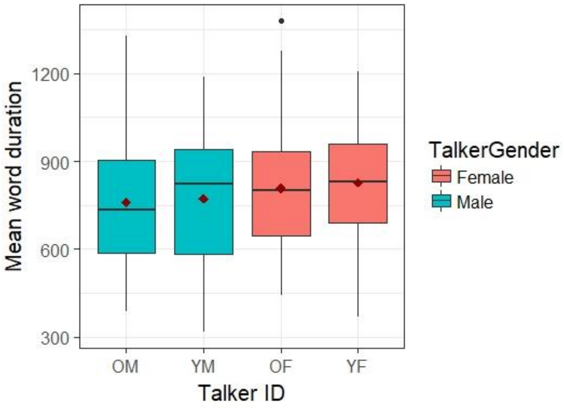


Figure 2.7. Variability in word duration by individual talkers: For each box, the upper and lower horizontal lines represent the first and third quartiles. The horizontal line and the point inside the box indicate the median and the mean. Data point represented by the whisker lie outside the middle 50%.

Evident in Figure 2.7 is that words tended to be produced with a shorter duration when the word was produced by the old male talker (OM). Mainly due to the idiosyncratic fast speech rate of OM, word durations appear to vary across gender and age of the talkers used in the experiment; words tended to be produced with shorter durations when spoken by male talkers than female talkers, and when spoken by older talkers than younger talkers.

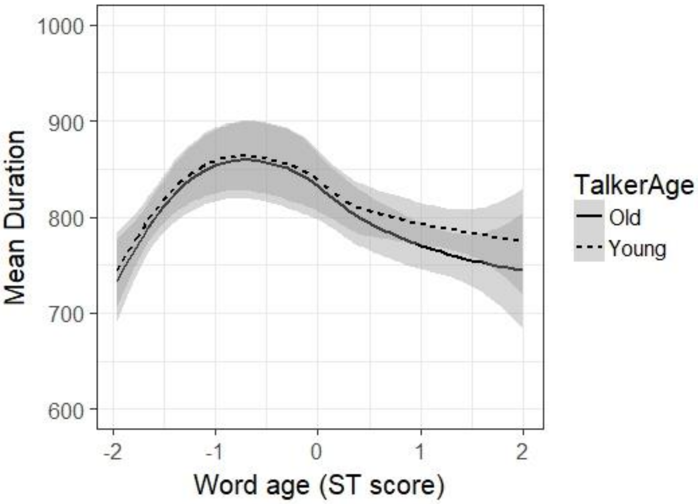


Figure 2.8. Mean duration of the critical items by word age and talker age: Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

Mean duration of the critical items are plotted by talker age and word age (ST score) in Figure 2.8. Durations of the items with word ages between -1 and 0.5 are longer than the rest of the items, in line with the RT patterns observed in both experiments (see Figure 2.6). As for an effect of talker age, the mean duration of older talkers is shorter than younger talkers regardless of word age in line with Figure 2.7. However, while the durational difference across talker age is marginal for items with word ages between -2 and 0, items with word ages greater than 0 exhibited a larger difference; as word age increased (i.e., the more associated with older people), younger talkers produced the word with longer duration than older talkers. These trends seem to indicate that, for the talkers in this experiment, production of old words was influenced by the predicted age-congruence lenitive effect, while production of young words was not.

To test whether production rate was slowed down by mismatch between talker age and word age, a frequentist linear mixed effects model was fit to raw word duration. Test variables were main effects and interaction of talker age and word age. Tested as fixed effects (through the procedure described in Section 2.1.2.1) were talker gender (binary, deviation coded), number of syllables (continuous), recording order of items (continuous), and their 2-way and 3-way interactions with the test variables. The fixed effects in the final model included main effects of talker gender and the number of syllables, as well as the test variables. Random effects included the by-item intercepts and a by-item slope for an interaction between talker age and word age.

Table 2.12. Summary of model fit to word duration

Model: lmer (Duration ~ TalkerAge*ST + TalkerGender + #Syllable + (1+TalkerAge:ST| Item))

	Estimate	Std. Error	t value	p value
(Intercept)	152.659	14.279	10.691	<.001
Talker age=young	16.042	4.435	3.617	<.001
Word age (ST score)	-3.965	2.870	-1.382	.167
Talker gender=male	-52.686	4.416	-11.932	<.001
#Syllable	237.050	5.103	46.454	0
Talker age=young : Word age	5.942	3.343	1.778	.075

As shown in the model's output (Table 2.12), word duration was significantly influenced by age and gender of the talkers; younger talkers produced words with longer duration than older talkers (*p*<.001), and male talkers produced words with shorter duration than female talkers (*p*<.001), in line with the patterns observed in Figures 2.7 and 2.8. The main effect of word age was not significant (*p*=.167), although words with higher word ages tended to be produced with shorter duration. Number of syllables also had an effect ($p<.001$); word duration was longer for words with a larger number of syllables. Along with these effects, the interaction between talker age and word age indicates that when the word was produced by younger talkers, word durations tended to be longer for words with higher word age; however, this effect did not reach significance ($p=.075$).

2.3.2. Filtering out age-related variability in word duration

This section reanalyzes the RT data from both experiments, using word duration residualized on an interaction between talker age and word age as a predictor for RT. First, similarly to the method used in Section 2.1.2.3, word duration was predicted by an interaction between talker age and word age in a linear regression model. The residuals from this model (i.e., residualized duration) constitutes a variable representing the durational variability, independent from the influence of the talker-word interaction. Then, residualized duration was included as a predictor for word duration, instead of raw duration, in the RT model for each experiment (i.e., the models reported in Table 2.4 and 2.8).

Table 2.13. Summary of RT model including residulalized duration (Experiment 1)

Model: lmer (RT ~ TalkerAge*ST + ResidualizedDuration + TalkerGender + TrialOrder +
(1+TalkerAge*ST | Participant) + (1+TalkerAge:ST+TalkerAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	1204.176	11.907	101.135	0
Talker age=young	-13.800	8.029	-1.719	.086
Word age (ST score)	21.981	4.454	4.935	<.001
Residualized duration	0.652	0.021	31.133	<.001
Talker gender=male	8.027	3.406	2.356	.018
Trial order	-0.030	0.014	-2.101	.036
Talker age=young : Word age	13.232	3.630	3.645	<.001

Table 2.14. Summary of RT model including residulalized duration (Experiment 2)

Model: same as in Table 2.13

	Estimate	Std. Error	t value	p value
(Intercept)	1208.878	17.674	68.397	0
Talker age=young	-30.952	6.382	-4.850	<.001
Word age (ST score)	21.774	4.900	4.444	<.001
Residualized duration	0.640	0.025	25.843	<.001
Talker gender=male	12.271	4.464	2.749	.006
Trial order	-0.040	0.010	-4.130	<.001
Talker age=young : Word age	11.737	4.342	2.703	.007

Table 2.13 and 2.14 present the outputs of the models fit to RT in Experiment 1 and Experiment 2, respectively. Both models show similar effects of all predictors as demonstrated in the RT models that included raw duration as a predictor (cf. Table 2.4 and 2.8). Particularly, an interaction between talker age and word age shows a significant effect in the same direction as in the previous RT models (*p*<.001 in Experiment 1, *p*<.01 in Experiment 2), indicating that, when listening to younger talkers, responses were slowed down as word age increased. These results

suggest that word recognition was influenced by age-indexed acoustics other than temporal properties in both experiments.

2.4. Discussion

Results from the two experiments reported in Chapter II demonstrate an influence of socio-indexical cues in the signal on recognition of socially-conditioned lexical items. In Experiment 1, word recognition facility improved when the word is produced by a talker from the age group that the word is associated with. Although the effect on accuracy is subtle, the effect is robust for recognition speed.

Experiment 2 replicates the age-congruence effect even when trials are not blocked by talker age. While the effect on RTs does not reach significance in the frequentist model, it is supported by the Bayesian model and by the effect size comparison showing that presentation mode does not play a significant role to observe the effect in the combined data. Moreover, the effect on accuracy rates is robust regardless of model type. Through the effect size comparison between the two experiments, I also demonstrated that no additional effect is driven by expectation about the talker prior to the target word onset. These results suggest that the effect does not rely on the listener’s expectations about talker age but could instead arise entirely from hearing words with the phonetic realizations they are most likely to be produced with.

I began Chapter I with a discussion about diverging viewpoints about influences of higher-level information on phonetic mapping. An important theoretical consensus between models with interative connections (e.g., TRACE) and bottom-up priority (e.g., Cohort), though, was that multiple words are activated by acoustic similarity and the activated candidates compete each other for recognition. The results from the two expeirments suggest that age-related information comes into play rapidly (i.e., with little resort to expectations) to help resolve the competition at some point. This finding challenges models that do not fully explain flexible and rapid adaption to talker variability, which instead assume that social information must be filtered out for roboust recognition. This assumption entails that social information is separable from phonetic information. However, the rapid integration of socially-indexed phonetic cues provides converging evidence for the sociophonetic framework, which has demonstrated that social information is so closely tied to phonetic cues (Johnson, 2005; Hay & Drager, 2007) that social cues are perceptually inseparable from phonetic processing (e.g., Van Berkum et al., 2008; Tesink et al., 2009).

However, the effects demonstrated in this chapter do not resolve the controversy on the timing of top-down feedback, nor can we pinpoint how "early" in the lexical processing the effect occurred, based on the loose linking hypothesis between the influence and the lexical decision reaction time. The observed effects can be accounted for by models supposing either interactive connections or a feed-forward system, with different predictions about the timing issue. While interactive models would allow an early feedback, proponents for bottom-up priority would argue that the expectation-free influence can also be interpreted to have occurred at a post-perceptual phase. In addition, as reviewed in Section 1.5.2, even an early influence is explained by bottom-up models that allow for probabilistic access processes or flexible changes of access routes based on the retrieved phonetic cues.

Bayesian models (Norris & McQueen, 2008; Kleinschmidt & Jaeger, 2015) do both, predicting that perception can be influenced by all sorts of prior probabilities, including listeners' prior knowledge about socially-conditioned distributions of words and phonetic realizations. Bayesian models in general would predict that hearing a talker from a certain age group builds up listeners' prediction and increases the probability of encountering a word that is likely to be used by that talker. More particularly, in Kleinschmidt and Jaeger's (2015) ideal adapter framework, listeners' beliefs about the cue distribution is concurrently updated and a talker-appropriate generative model is selected by the incremental evidence observed from the signal. Thus, when in a novel situation (e.g., young words produced by an older talker), uncertain beliefs about the underlying distribution lead to an adaptation process and result in slower recognition, while frequently experienced cue distributions (e.g., young words produced by a younger talker) are recognized immediately.

The effect is also predicted by exemplar models of speech perception (Johnson, 1997; Pierrehumbert, 2001, 2002), in which phonetically-rich representations of words are stored in memory, and lexical access is determined (partially) by similarity between acoustic detail in the signal and that in the lexical representation. Under this mechanism, age-linked phonetic cues in the signal automatically activate relevant age-linked word-based memories. As a result, recognition is facilitated when the signal closely matches the representation.

In Experiment 1, I also demonstrated that the effect is predicted by both usage- and stereotype-based word age. In addition, through a series of model comparisons (2.1.2.3), I showed that stereotypical associations between words and age groups appear to wield an additional influence above and beyond the effect of distribution, at least when listeners are exposed to words on a wide spectrum of age-related variability. Although it may be questionable whether individuals' exposure frequency is efficiently captured by the self-reported usage frequency that I used (or even corpus-based frequency), this finding seems to suggest a possibility that activation of word-based memories receives an "added boost" when higher-level (stereotype-based) social information is involved.

This latter effect may not be completely explained by purely episodic models in which statistical distributions of phonetic realizations play a pivotal role. However, the hybrid approach to exemplar modeling (Pitt, 2009; Pinnow & Connine, 2014; Pierrehumbert, 2003, 2016) puts forward a mechanism in which perception can be affected either by specific exemplars or by abstract social information at a higher level. In my experiments, experience with specific realizations may be generalized to form abstract indices based on indexical features, such as contextual domains (e.g., internet slangs or traditional kinship terms) or social changes (e.g., new concepts or antiquated terms). And, the belief that certain words are skewed in distribution is also refined by native speakers' conscious awareness of, or metalinguistic discussions about, socially-conditioned variations (Labov, 1972; see also Drager & Kirtley, 2016, for different degrees of such beliefs and their effects on speech perception).

These results, taken together with previous results such as those outlined in Section 1.3 and 1.4, raise an empirical question whether stereotypes can in fact override (i.e., strengthen or even change the direction of) the effect of pure frequency. If so, it would suggest that lexical representations are formed and shaped by experience and distributional properties but the social indices of the representations are evaluated and reinforced by social stereotypes about words. In this vein, models of speech perception should take into account the interaction between the distributional factor and the ideological distinctions of social groups, which is in no way negligible amid the formation of cognitive links used in linguistic processing.

In addition, the analyses for effects of age-congruence on word durations (Section 2.3) point to some implications for realizations of subphonemic detail. Although the effect did not reach significance – possibly because the talkers were instructed to maintain the rhythmic patterns consistent across items – words tended to be produced with a shorter duration when the talker and the word matched in age. To verify Hay and Foulkes's (2016) prediction that word-specific production details are conditioned by the word's distribution across social categories, production of age-indexed words may be reexamined with a larger number of speakers in the future, integrating other acoustic parameters for phonetic reduction (e.g., vowel formant values, VOT, gestural kinematics). If such evidence is established from a sociophonetic perspective, it would also lend more general support for usage-based phonetic grammar, which predicts, for example, that frequently activated words are produced with lenitive forms (Bybee, 2000; 2001; 2002).

So far, this chapter has highlighted the role of socially-indexed phonetic detail in lexical access. Assuming that the age-congruence effect I demonstrated in this chapter could have been triggered by a vast range of age-indexed phonetic cues, a question that ensues is to what degree distribution of phonetic realizations of a particular variable can be linked to the socially-conditioned lexical items. The experiment reported in Chapter III is going to further explore the nature of the age-congruence effect along this question.

CHAPTER III

The Link between Lexical Items and Phonetic Variants

In Chapter II, I provided evidence that lexical access benefits from socio-indexical phonetic details when the lexical item is also indexed to the congruent social category. In the line of research demonstrating an age-congruence effect on lexical access, including Walker and Hay (2011) and the experiments reported in Chapter II, age-related phonetic information is contrasted by having talkers of different ages produce the entire word. This means that listeners in those experiments were likely provided with multiple phonetic cues indexing talkers' social information.

This chapter further explores the socially-conditioned links between lexical items and phonetic realizations using a phonetic priming experiment. As reviewed in Chapter I, it is well known that even fine-grained phonetic details are linked with socio-indexical information, and listeners are able to exploit these links during speech perception. Thus, semantically related lexical items are robustly primed when a single phonetic variant contained in the prime word is manipulated to resemble the existing mental representation of individual listeners (Sumner & Samuel, 2005, 2009).

There is still much to be learned about the degree to which phonetic- and lexical-level variations are linked in memory and are utilized in lexical access. Since sociophonetic variation takes place under a highly systematic structure, in which phonetic forms and lexical items often covary in accordance with social categories, it is conceivable that occurrence of a socially-conditioned phonetic variant may also be used to guide lexical access to lexical items that are associated with the congruent social information.

In Chapter III, I present an experiment testing whether lexical access can be primed by talker age cues that are restricted to a single sociophonetic variable and are presented prior to the target word. Such a phonetic priming effect, if observed, would suggest a close link between phonetic variants and lexical items, and highlight the role of probabilistic inference in lexical processing. The results will be discussed in the context of cognition models, focusing on whether activation of abstract social information at a higher level is required in the effect, or whether the effect is such a highly automatic process that lexical access can be directly guided by a socio-indexical phonetic variant via inter-unit connection.

3.1. Experiment 3: Lexical decision task preceded by age-associated priming

In Experiment 3, the lexical decision paradigm is combined with a phonetic priming method. Each trial consists of a lexical decision task preceded by a color-identification task that serves as a priming stage. Listeners are exposed to a prime word that contains an acoustically manipulated variable, mimicking either older or younger Korean speakers' phonetic realizations of the variable. Then, the effects of the priming on lexical access are evaluated by accuracy and reaction times in the lexical decision task. Since the manipulations are made only in the prime word (not in the target word itself), it allows me to test whether a previously encountered socio-indexical variant can affect subsequent processing of an age-related target word produced by the same talker.

3.2. Ongoing sound change in Korean plosives

As an age-related phonetic variable, I used phrase-initial plosives, which are undergoing sound change in contemporary Seoul Korean. This section discusses the sound change. In Section 3.2.1, I will briefly introduce the acoustic properties found in the typologically rare, three-way laryngeal contrast in Korean stop. Next, I will step through an overview of the sound change, providing evidence from production (3.2.2) and perception studies (3.2.3). Finally, 3.2.4 explains the motivation of the sound change from a viewpoint that the sound change in Korean stops is led by a tonogenetic process in Korean phonology.

3.2.1. Acoustic properties of Korean plosives

Plosives in Korean are all phonetically voiceless but three phonemic categories are distinguished by phonation type: lax, aspirated, and tense. Affricates and fricatives in the alveolo-palatal position are also differentiated by the laryngeal settings; affricates are also grouped into three categories and fricatives have a tense-lax contrast. The three-way categorization of Korean obstruents is summarized in Table 3.1.

Table 3.1. Three-way distinction of Korean obstruents by place and manner of articulation[22]

Manner		Plosive			Affricate	Fricative
Place		Bilabial	Alveolar	Velar	Alveolo-palatal	
Phonation type	Lax	p (ㅂ)	t (ㄷ)	k (ㄱ)	tɕ (ㅈ)	s (ㅅ)
	Aspirated	pʰ (ㅍ)	tʰ (ㅌ)	kʰ (ㅋ)	tɕʰ (ㅊ)	–
	Tense	p* (ㅃ)	t* (ㄸ)	k*(ㄲ)	tɕ* (ㅉ)	s* (ㅆ)

Normally, three parameters are used to describe their acoustic properties. First, they are primarily distinguished by voice onset time (VOT). According to Jun's (2006) summary of production studies conducted in 1990s, tense has the shortest mean VOT (10ms), aspirated is the longest (106ms), and lax is in between (39ms).

Second, because the three types are produced with different degrees of muscle tenseness and glottal pressure, they are also distinguished by voice quality of the following vowel, which is acoustically captured by amplitude difference between the first and second harmonics (H1-H2). Lax plosives have the greatest H1-H2 difference in the following vowel (resulting in breathy voice), tense plosives have the smallest (pressed voice quality), and aspirated plosives are in between (Cho et al. 2002; Kim, 2008; Kong, Beckman, & Edwards, 2011).

Last, muscle tenseness also affects fundamental frequency (F0) at voice onset after the burst. Everything else being equal, tense is produced with a higher F0 value than lax, but tense

---

[22] The categorization of Korean obstruents follows Cho et al. (2002) and Shin & Cha (2011). They all agree that the alveolo-palatal obstruents, i.e., 'ㅈ, ㅊ, ㅉ' in Hangul, are affricates while some authors categorize them as stops. Phonetic transcriptions for the affricates also vary in literature, but /tɕ/ is used following Shin and Cha (2011). Hangul letters are accompanied in parentheses to minimize confusion.

and aspirated plosives show a wide overlap (Dart 1987; Cho et al. 2002; Kim 2004; Kong et al. 2011; Chang 2012).

3.2.2. Change in acoustic properties

An ongoing sound change in the stop system is reported in a number of apparent-time production studies on Seoul Korean (Silva, 2002, 2006; Wright, 2007; Kang & Guion, 2008; Oh, 2011; Kim, 2012, 2013, 2014; Kang, 2014). For example, using corpus data of 117 Seoul Korean speakers (recorded in 2003, age: 19-71), Kang (2014) demonstrates that stops found in the phrase-initial position are realized with different VOT and F0 at the voice onset between age groups. Specifically, younger speakers tend to produce aspirated stops with shorter VOT (thus overlapping with lax stops), but with higher pitch (as compensation of the VOT merger). In contrast, older speakers tend to maintain a VOT-based distinction. Additionally, tense stops – which already have the greatest phonetic motivation for higher pitch among the three types – are produced with further raised F0 in the innovative realizations.

Lax stops, on the other hand, do not seem to be undergoing substantial changes in either VOT or F0 values. Although a handful of studies in the 1990s found that the VOT merger is induced by VOT changes in both aspirated and lax plosives (Silva, 1992; Cho, 1996; Han, 1996), no cross-age difference in the VOT-F0 dimension is confirmed in more recent research. Cross-age differences in VOT/F0 values of the three categories are summarized in Table 3.2.

Table 3.2. Overview of VOT and F0 found in younger/older Seoul speakers' realizations of plosives: Each phonation type is paired with a prime word used in Experiment 3 (see 3.4.2.2 for detail).

Phonation type	VOT	F0 at voice onset	Prime word
Lax	no difference	no difference	[pamsɛk] 'brown'
Aspirated	**old > young**	**old < young**	[pʰaɾaŋ] 'blue'
Tense	no difference	**old < young**	[p*alkaŋ] 'red'

3.2.3. Change in cue-weighting

The production studies outlined above suggest that the traditional VOT-based distinction between lax and aspirated sounds is being gradually replaced by an F0 distinction, which had been a secondary feature in the past. Not only is this shift in cue-weighting found in production data, but there is strong evidence that the perception of younger listeners has also shifted in such a way that F0 has become a primary perceptual cue as a consequence of probabilistic inference for cue-weighting.

Studies in the early 2000s began to reveal that younger listeners (age: 20-32) use F0, as well as VOT, as a reliable cue to differentiate aspirated and lax stops (Kim, Beddor, & Horrocks, 2002; Kim, 2004). In recent research, Schertz, Kang, and Han (2016) tested listeners using conservative dialects (heritage dialects spoken in Hunchun and Dandong, China), and found that whether VOT or F0 is a more reliable cue for the stop categories depends on either the talker's or the listener's age. In other words, F0 tends to be weighted greater than VOT when the listener is younger, or when the talker is younger.

Schertz et al.'s results also indicate that the cue-weighting shift in perception may have spread out across dialects. Speakers in Kyeongsang Province, in contrast with speakers from Seoul, are primarily found to use VOT to differentiate both in production (Lee & Jongman, 2012) and perception (Lee, Politzer-Ahles, & Jongman, 2013). However, this is because the Kyeongsang Dialect is a tonal dialect where F0 serves as a cue to lexical pitch accent, rather than a cue to the laryngeal type[23].

Gender also appears to wield an influence on cue distribution. Since the sound change has been led by younger female speakers (Oh, 2011; Kang, 2014) – a population that often leads phonetic changes (Labov, 1990) – the F0 cue is highly weighted in stop realizations of female speakers (Kang, 2014), and in the perception of tokens produced by females (Kong et al., 2011).

Taken together, these findings suggest that the statistical distribution of socio-indexical cues are stored in memory based on individuals' prior experience, and listeners use the link between social information and phonetic cues to achieve robust perception of the variable. The findings are consistent with experience-based models positing bidirectional interplay between linguistic and social information (Johnson, 1997; Sumner et al., 2014; Kleinschmidt & Jaeger, 2015).

3.2.4. Change in phonological structure

The previous sections provided background of an ongoing sound change in Korean stops, focusing especially on the increased role of F0 as a cue to segmental categories; Korean stops are going through a systematic shift in cue-weighting from VOT to F0 both in production and

---

[23] This is notable for the current study, as there were five Kyeongsang speakers who participated in Experiment 3 (see Section 3.4.5). However, no difference in behavior of these participants were found.

perception in at least some dialects. As one of the most compelling account for its motivation, some scholars regard the change as part of a tonogenetic process in the Korean phonology (e.g., Silva, 2006; Wright, 2007; Kang, 2014)[24]. In general, voiceless consonants in initial position that are contrasted by consonantal features, such as voicing, aspiration, phonation types, tend to be realized with higher pitch than voiced ones (as a redundant phonetic byproduct) due to high aerodynamic pressure across the vocal folds (Hombert, Ohala, & Ewen, 1979)[25]. As tonogenesis proceeds, the consonants go through more and more pitch perturbation, and eventually tones become the primary cue to the contrast, replacing consonantal features.

The tonogenetic approach to the sound change in Korean stops is bolstered by Kang's (2014) finding that the enhancement of F0 cue weighting extends to influence *all* [+spread glottis] consonants. She argues that the cue-weighting shift is not just taking place as a compensatory enhancement targeting the aspirated-lax contrast which lost VOT contrast, but is a consonant-induced structural change that overarches the intonational manifestation of Korean. This observation is in line with the Accentual Phrase (AP) structure of Seoul Korean (Jun, 1993, 1996; 2011), in which an AP that begins with a [+stiff vocal folds] consonant is realized with H+H boundary tones while an AP with initial lax stops or sonorants are realized with L+H.

---

[24] Tonogenesis refers to a phonological change through which a non-tonal language acquires tonal contrasts, which is well documented for Vietnamese, Chinese and other Southeast Asian languages (Maran, 1973; Matisoff, 1973; Hombert et al., 1979).

[25] For the same reason, VOT and F0 are positively correlated in many other languages including English and French (Kohler, 1982, 1984, 1985; Kingston & Diehl, 1994; Torre & Barlow, 2009). In addition, there is evidence that English listeners track F0 cues to discriminate voicing contrasts in the absence of VOT cues (Castleman & Diehl, 1996; Idemaru & Holt, 2011).

3.3. Predictions

With age-associated covariation between the plosives and lexical items introduced in Chapter II, Experiment 3 tests whether exposure to a stop variable manipulated across different conditions can influence subsequent lexical access. The current experiment tests the following two hypotheses:

First, it is expected that listeners' expectations about the talker are guided by phonetic realizations of the stop variable, provided that that variation is the only cue to the talker's age. Because aspirated and tense stops are produced differently by younger and older speakers (while lax stops are not), the VOT and F0 contained in these sounds are manipulated in the current experiment (see 3.4.2.2). I predict that lexical access will be facilitated when the word is preceded by a prime word containing an aspirated or tense stop (ASP/TENSE, henceforth) that indexes the age group that the word is also associated with, suggesting that socially-conditioned phonetic variants and lexical items are closely linked in memory and listeners use the link to guide lexical access.

Second, by including trials preceded by the invariant lax stop (LAX, henceforth), I explore a secondary research question: whether exposure to the age-related ASP/TENSE variants produced by a talker can activate that talker's age at a higher level, and then be generalized to trials with the unvarying LAX prime word produced by that talker. If activation of abstract social information is necessary to induce the age-congruence effect of ASP/TENSE, recognition of words preceded by LAX may also be affected by whether the talker in the given trial is one who produces old- or young-associated variants of ASP/TENSE. On the other hand, if phonetic variants are directly indexed to lexical items that are associated with the same age group, lexical access in LAX trials may not be affected by individual talkers, no matter which age group the talker's realizations of ASP/TENSE are associated with.

3.4. Method

3.4.1. Talkers

Four talkers were selected based on a norming test, in which short clips of speech produced by 14 speakers (8 males and 6 females, ages between 26 and 44) were rated by 40 college students (participants in other experiments) for voice characteristics including dialect, likelihood of using language associated with young people, and perceived age. The select talkers were ones who were heard as Seoul Dialect speakers, who were neutral in terms of the likelihood rating, and whose biological ages and perceived ages were close to 40, a neutral age population in terms of association with the age-associated lexical items.

Half of the trials including all critical trials (see Section 3.4.2.3 for composition of stimuli) were produced by two female talkers (age: 40 and 41) and their realizations of the prime words were acoustically manipulated (see 3.4.2.2). In addition, two male talkers (age: 34 and 36) were used as distracters to divert listeners' attention from the manipulation, and to prevent listeners from noticing that only certain sounds were used as the initial sound in the target words (see 3.4.2.3 for detail).

Female voices, rather than male voices, were manipulated for two reasons. First, females' speech in general utilizes a wider range of pitch contour than males, and so the priming stimuli would sound more natural when F0 is manipulated to change within a bisyllabic word (see 3.4.2.2 for detail). Second, the VOT merger is almost complete in younger females' speech but is less advanced for older males (Kang & Han, 2013; Kang, 2014). Thus, using male talkers would decrease the probabilities for listeners to use F0 cues (Kong et al., 2011)[26].

3.4.2. Stimuli

3.4.2.1. Recording. The talkers were naïve as to the research purposes at the time of recording. Their productions of priming stimuli and lexical decision stimuli were recorded in quiet rooms, using the same equipment and method described in 2.1.1.4. Corrections were made when mispronunciation occurred. For the priming stimuli, talkers were instructed to produce the color terms in an H+L% prosodic structure, and no instructions about VOT or aspiration were given.

3.4.2.2. Priming stimuli. Priming stimuli were created using a matched-guise design (Lambert et al., 1960), which has been widely adapted in sociolinguistics to examine listeners' attitudes or impressions about talkers using particular variants. In experiments using this method with acoustic manipulation (e.g., Campbell-Kibler, 2007; Yi, 2015), a single recording is taken from a talker and edited into multiple guises for that talker by splicing just the manipulated portion into the raw token. This allows the researcher to elicit listeners' reaction to the tested variable independent from traits associated with other auditory cues found in the stimuli.

Three color terms – *brown* (LAX, [pamsɛk]), *blue* (ASP, [pʰaɾaŋ]), and *red* (TENSE, [p*alkaŋ]) – were used as priming words, each of which begins with one of the three phonation types of the Korean bilabial stops, followed by the vowel [a]. VOT and F0 in ASP and TENSE

---

[26] Note, however, that female voices generally undergo a lowering of F0 with ageing, while F0 of male voices raises (Mueller, Sweeny, & Baribeau, 1984; Honjo & Isshiki, 1980; Torre & Barlow, 2009). Since the sound change at hand accompanies F0 raising in younger speakers' speech, replicating this experiment with male voices would be worthwhile for future work.

tokens produced by the female talkers were manipulated to create two guises (older and younger) per talker, while a single LAX token was used (with no manipulation) per talker. Therefore, a total of eight manipulated priming stimuli were created (2 plosives x 2 guises x 2 talkers) plus two LAX tokens (one per talker). VOT/F0 values of the original recordings and manipulated tokens are presented in Table 3.3, and the priming stimuli are plotted on a VOT-F0 dimension in Figure 3.1.

Table 3.3. VOT and F0 values of the original recordings and manipulated tokens: VOT values are represented in ms, and F0s are in Hz.

Talker ID	Prime ID	Original		Manipulated variable	Younger guise		Older guise	
		VOT	F0		VOT	F0	VOT	F0
Talker A	LAX	41	209.9	*Single unmanipulated token used*				
	ASP	80	276.8	VOT & F0	39	325.8	93	210.3
	TENSE	6	256.5	F0	6	316.1	6	210.4
Talker B	LAX	38	204.4	*Single unmanipulated token used*				
	ASP	54	267.0	VOT & F0	39	317.1	94	205.3
	TENSE	10	242.1	F0	10	308.3	10	206.1

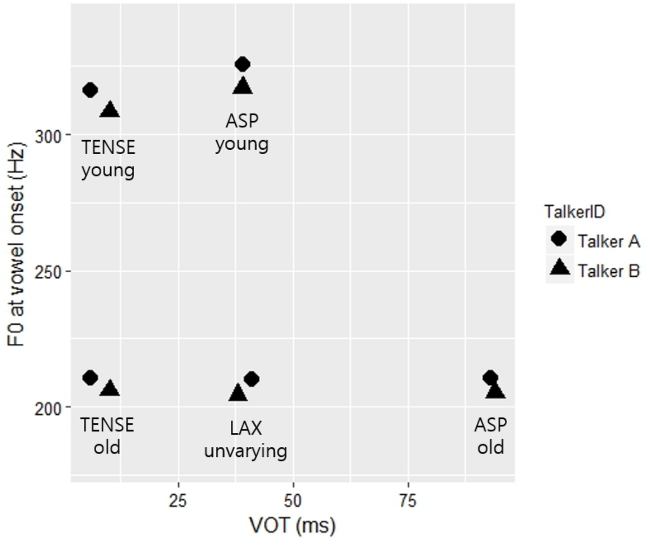


Figure 3.1. VOT/F0 distribution of priming stimuli

Recall that the acoustic change in Korean stops is parameterized by a cue-weighting shift from VOT to F0, and that perception is adjusted to the speaker's cue distribution. Accordingly, the acoustic manipulation was made in such a way that the two guises were contrasted by different cue-weighting strategies; the older guise relied on *only* VOT, and the younger guise *maximized* the use of F0 cue[27]. Due to the exclusive reliance on VOT for the older guise, extreme combinations of VOT and F0 are implemented for ASP and TENSE tokens. But, all values of the manipulated tokens fall within the VOT/F0 range found in the production literature (e.g., Kang, 2014), so that responses can be drawn by either (1) the phonetic form itself, regardless of talker ID (e.g., via episodic memory trace), or (2) whether VOT or F0 is weighted greater in individual talkers' cue distribution (via abstract social information).

It should be also noted that F0 values found in the literature vary considerably. Even in older speakers' speech, aspirated and tense stops tend to be realized with higher F0 than lax stops due to consonant-induced F0 perturbation (see Section 3.2.4). In Kang (2014), however, the mean difference in F0 between aspirated/tense and lax tokens gradually decreases as the age of the speaker increases from 19 to 71, and is almost non-existent in the speech of the six oldest speakers (born in the 1930s). Thus, stimuli in the older-guise (which have little F0 difference) are representative of stops produced by older people. The stimuli are available for download from: https://www.bloomsbury.com/cw/experimental-research-methods-in-sociolinguistics/.

The acoustic manipulation was made using Praat. VOT values (reported in Table 3.3) were measured from the plosive release to the vowel onset. The release point was determined at the point in which a sudden increase of amplitude is observed in the waveform at the word onset.

---

[27] The inclusion of LAX also provides the listeners with a reference point for within-talker distribution of VOT and F0.

To define the vowel onset, the starting point of periodic cycles was determined in the waveform, and the onset of vowel formants was also compared using the spectrogram. For VOT manipulation, periodic zero-crossing cycles were taken from the original token's VOT region, and then these cycles were either removed to create tokens with shorter VOT or spliced to create tokens with longer VOT.

F0 values were measured automatically using Praat, at 20ms past the point at which an F0 value was definable. For F0 manipulation, pitch contours were manipulated using the "To Manipulation..." function in Praat. First, the pitch curve in the original recording was stylized by semitones. This automatically created pitch points outlining the original pitch curve. Using the mouse, the first pitch point was adjusted to the desired vowel-onset F0 value. For example, the first pitch point was dragged down to create an older guise token of ASP and TENSE, or raised to create a younger guise token. Then, subsequent pitch points were adjusted to meet the offset of the first syllable in such a way that the manipulated pitch contour parallels with the original contour in the early region and gradually merges with it from the midpoint of the first vowel. Additional pitch points were added manually when necessary. The manipulated contour completely merged with the pitch contour of the original token at the endpoint of the first syllable, and the pitch contour in the second syllable was reconstructed by adding more pitch points along the original pitch curve. As a result, only the pitch on the first syllable was audibly different from the original token. Example pitch contours of the priming stimuli are provided in Figure 3.2.

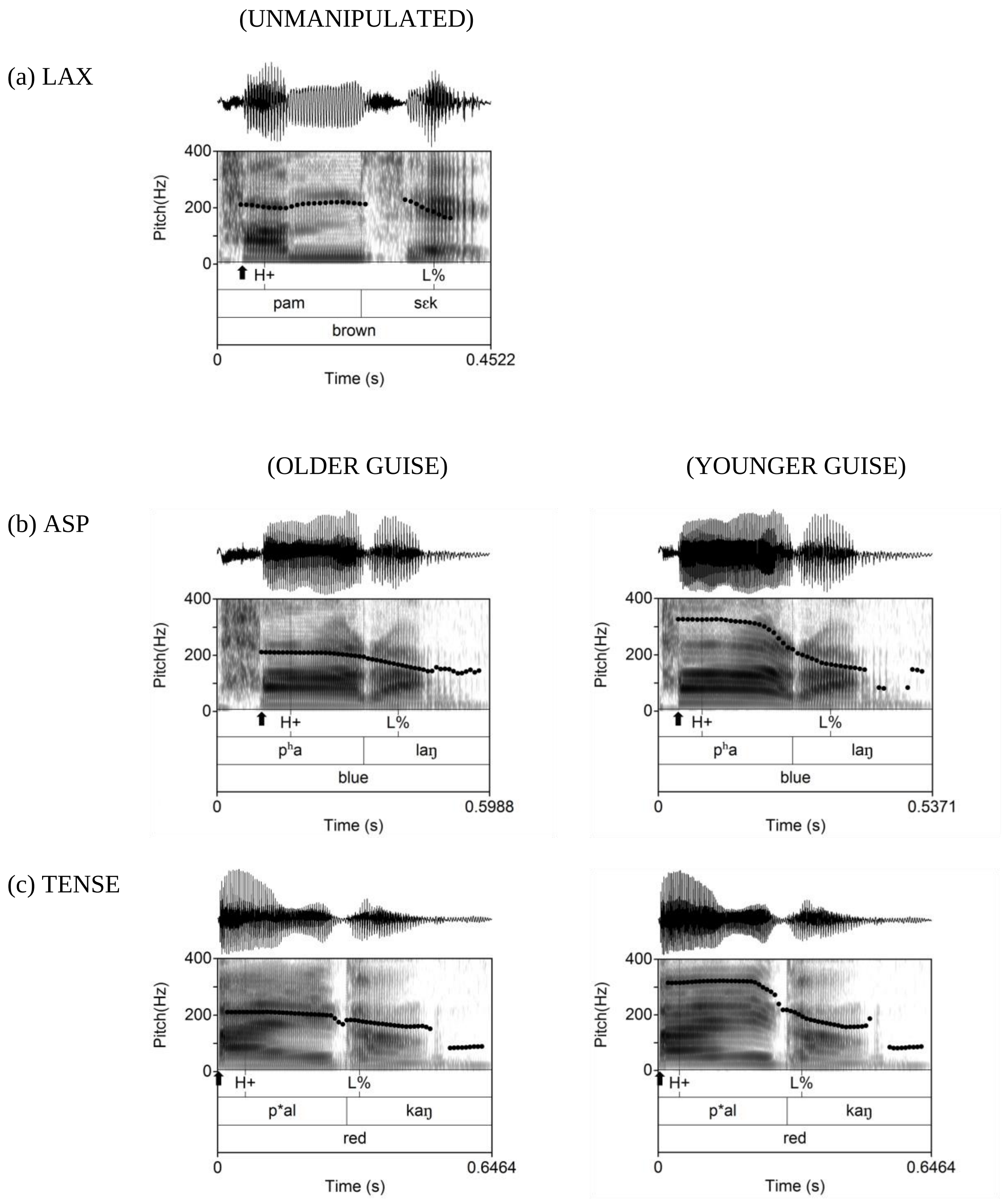


Figure 3.2. Example pitch contours of priming stimuli produced by a female talker: (a) LAX [pamsɛk], "brown", (b) ASP [p^{h}aɾaŋ], "blue", (c) TENSE [p*alkaŋ], "red". In each figure, voice onset is indicated by an arrow underneath the spectrogram.

In Figure 3.2, the pitch contours of tokens in the younger guise show a clear H+L% pattern, with a higher pitch at the voice onset than for tokens in the older guise. Although the pitch difference between syllables in the older guise may be subtle in the pitch curve, an H+L% pattern is also clearly audible in the auditory stimuli. Also apparent in Figure 3.2 is that VOT of LAX tokens (duration between Time=0 and the arrow) is comparable with that of the ASP tokens in the younger guise, mimicking the VOT merger in productions of younger speakers, whereas ASP tokens in the older guise have a longer VOT than LAX tokens as well as ASP tokens in the younger guise.

3.4.2.3. Lexical decision stimuli. Lexical decision stimuli were composed of 270 real words and 270 non-words chosen from those used in the experiments reported in Chapter II (see Appendix I for a list of real-word items). A single auditory token was used for each lexical decision item across guises within each female talker. In order to limit the occurrence of the female talkers' age-related cue realizations to only the priming words, items that begin with plosives, affricates, and the two fricatives (s, s*) were restricted to the male talkers (N=270). Items with initial glottal fricative /h/ – which is produced with the least oral constriction among obstruents, and is not contrasted by phonation type or closure release – were included in the female-talker items to maximize the number of observations. As a result, the female talkers produced 270 real-word and non-word items that begin with vowels (N=75), glides (N=50), nasals (N=81), liquids (N=7), and /h/ (N=57).

Table 3.4. Summary of lexical decision stimuli

All items	Lexical status	Produced by	Word onset type	Word age (mean duration)
N=540	Real words (N=270)	Female talkers (N=146)	Non-obstruents and /h/	**37 old (624ms),** **39 neutral (634ms),** **34 young (627ms)**, 36 fillers
		Male talkers (N=124)	Obstruents	31 old, 31 neutral, 31 young, 31 fillers
	Non-words (N=270)	Female talkers (N=124)	Non-obstruents and /h/	NA
		Male talkers (N=146)	Obstruents	NA

Table 3.4 provides a summary of the lexical decision items. Unbalanced numbers of real words and non-words were used across the male and female talkers in order to maximize the number of real-word observations from the female-talker trials; for the female talkers, a greater number of real words (N=146) were assigned than non-words (N=124), whereas for male talkers a greater number of non-words (N=146) were assigned than real words (N=124).

Table 3.4 also provides a summary of the real word items based on the categorical distinction of word age used in 2.1.1.2, indicating that the items are evenly distributed over word age. However, word age will be treated as a continuous variable in the analysis (Section 3.5). Fillers are the top-frequency words in the Sejong Corpus, and are not assigned a word age. Among the real words produced by the female talkers, 110 items that had been assigned word age were used as critical items for the analysis. In Table 3.4, these items are marked in bold and provided with mean word duration for each word age group.

3.4.3. Procedure

The experiment was run in a quiet room using the E-Prime 2.0 software on three different laptop computers. Participants were seated in front of a response pad (model: Cedrus RB-530, or RB-834), which had three buttons for color identification on one side and two buttons for lexical decision on the other side (see Figure 3.3). Participants alternated between two tasks in each trial. First, a priming stimulus (a color term produced in isolation) was played over the headphones and participants were instructed to press a button with their recessive hand, indicating what color they heard. Then, they heard a lexical decision stimulus and pressed a "Real" or "False" button with their dominant hand, indicating whether they heard a real word.

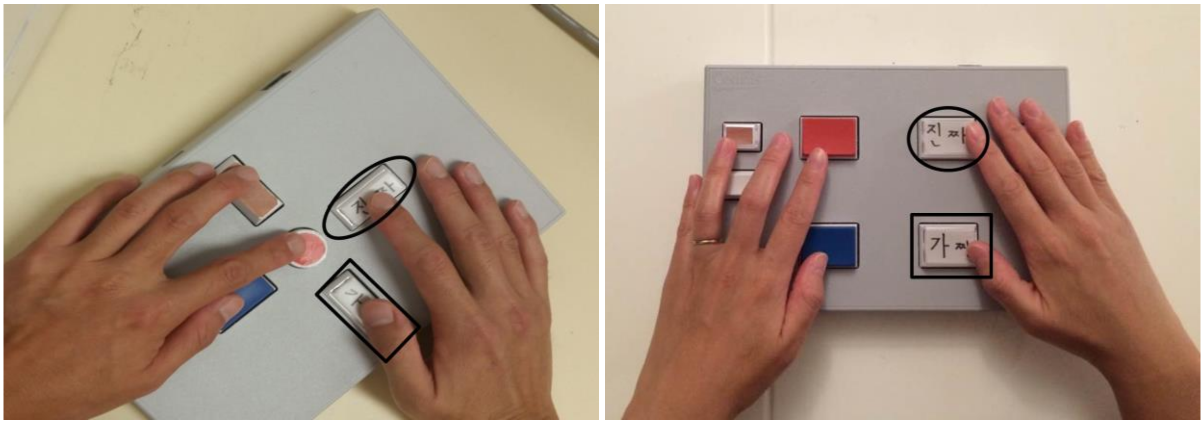


Figure 3.3. Photos of the two response pads with key-top settings for right-handed participants: Participants responded with their non-dominant hand to the primes and with their dominant hand to the targets.

Figure 3.3 shows the response pads with key-top settings arranged for right-handed participants. Two different response pads were used by different participants, so that more than one participant could be run at a time. Right-handers were randomly assigned one of the two response pads. Left-handers were assigned the one in the left panel; for these participants, the orientation of the pad was turned upside down and the key-tops were replaced before the experiment session began, so that all participants responded to the tasks with designated hands.

Fingers were also designated for pressing each of the five buttons. Participants pressed "Real" with the index finger (top-right, marked by a circle in each photo) and "False" with the thumb (bottom-right, marked by a rectangle) of their dominant hand. With the other hand, they pressed red with the index finger, blue with the thumb, and brown with either the middle or the ring finger. Although the button configuration was different between response pads, the fingertip positions were almost identical, by twisting the orientation of the response pad in the left panel by 45 degrees to the counterclockwise direction. No differences in participant behavior were either hypothesized or observed across response pads and across handedness.

Participants were also informed that they were going to hear two male and two female voices, in order to help them track the cue distribution of each talker. After every 135 trials (1/4 of the session), participants were able to choose to take a break (while remaining seated). Except for the break, they were instructed not to move their hand off the response pad.

An exit survey was conducted after the experiment session. First, participants were asked what the purpose of the experiment might be. Then, perceived age of each talker was measured. Participants heard the priming stimuli again (in the order of ASP, TENSE, and LAX, realized in the guise that each participant was exposed to) and were asked to guess the age of each talker. They were instructed to provide a specific number (e.g., 19, 20, 21), as opposed to an age range (e.g., 20s, about 20, 19-21).

3.4.4. Design

The priming stimuli were presented in a between-subject design; for example, half of the participants heard female talker A in the younger guise and female talker B in the older guise, and vice versa for the other half of the participants. Lexical decision items were counterbalanced by talker ID (female talker A and B) and guise (older and younger) into four lists in a Latin Square design. Each participant was randomly assigned to one of the four item lists, and items appeared in a random order for each participant. Each lexical decision item was paired with one of the three color terms, and color terms were not counterbalanced across items.

3.4.5. Participants

40 Korean-native college students (27 females and 13 males, age: 18-27) took part in the experiment at two locations: 20 at Chung-Ang University in Seoul, Korea and 20 at the University of Hawaiʻi at Mānoa. All Hawaiʻi participants had been in Hawaiʻi for less than two years at the time of participation.

All but five participants listed the Seoul Dialect as their most frequently used dialect. These five participants (3 males and 2 females) listed the Kyeongsang Dialect as primary and the Seoul Dialect as secondary, but they all reported they are also fluent in the Seoul Dialect.

3.5. Results

3.5.1. Accuracy

The overall accuracy rate of all real word trials (N=10,800) was 87.16%. This is higher than Experiment 1 (86.98%) and 2 (84.64%) despite the addition of the color identification task, which might be expected to decrease accuracy by involving the participants in unrelated processing.

Among the 4,400 data points collected from the 110 critical items by 40 participants, trials with an inaccurate color press (N=58) were excluded for both accuracy and RT results, because preliminary analysis revealed that lexical decision response was significantly worse in accuracy and RT when listeners pressed an incorrect color button, an effect I attribute to a difference in attention. Excluding them, 4,342 tokens were analyzed for accuracy in the lexical decision task. The accuracy rate of this dataset was 83.00%[28]. In Figure 3.4, accuracy rates for the critical items preceded by ASP/TENSE (dataset (a)) and LAX (dataset (b)) are plotted by word age and guise, respectively.

---

[28] Accuracy rate is reduced compared to all real word trials because trials with filler items (i.e., words with high frequency) are removed.

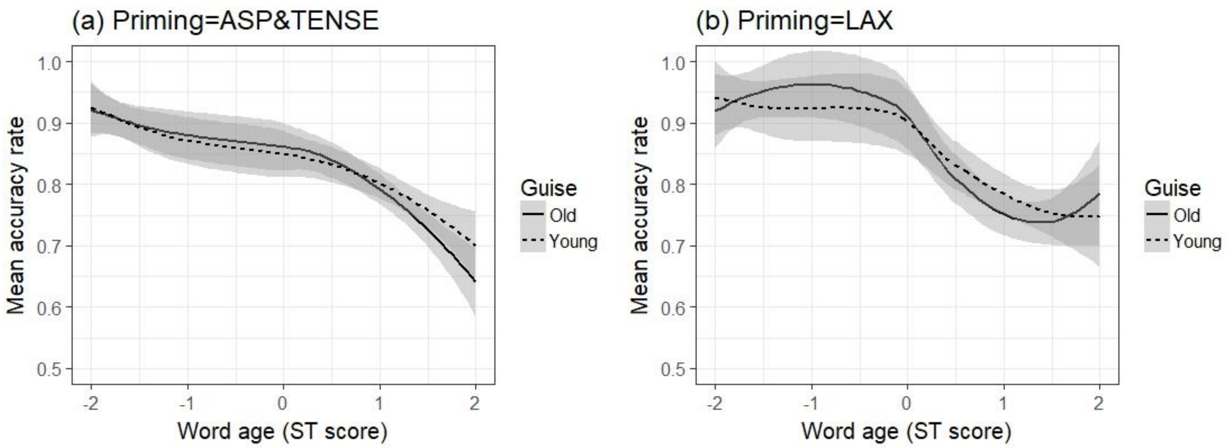


Figure 3.4. Mean accuracy rates by word age and guise: Trials preceded by (a) ASP/TENSE (N=2,914) and (b) LAX (N=1,428). Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

The predicted pattern is not observed in either graph; recognition accuracy was not improved when the word was preceded by a stop variable realized in a guise that matched word age. A binomial mixed effects model was fit to accuracy of the dataset (a). Fixed effects and random effects were tested using the same method in 2.1.2.1. Fixed effects in the final model included word age, guise (deviation coded), and their interaction as test variables, and Prime ID (ASP or TENSE, deviation coded) as a control variable. Included as random effects were a by-participant slope for word age and intercepts for by-participant and by-item variance. In this model, no significant interaction was found between guise and word age (β=0.070, s.e.=0.091, *p*=.444). A separate model was fit to the dataset (b), with the same fixed effects as above, along with by-participant and by-item random intercepts. Guise and word age did not interact in this model, either (β=0.127, s.e.=0.145, *p*=.382).

3.5.2. Reaction times

RT tokens for the lexical decision task were obtained from real words produced by all talkers to which both tasks were responded correctly (N=9,283). Among them, 159 outliers were removed (<2%) through the same process described in 2.1.2.2. Figure 3.5 compares the distribution of RTs after outlier treatment, between Experiment 3 and the two previous experiments.

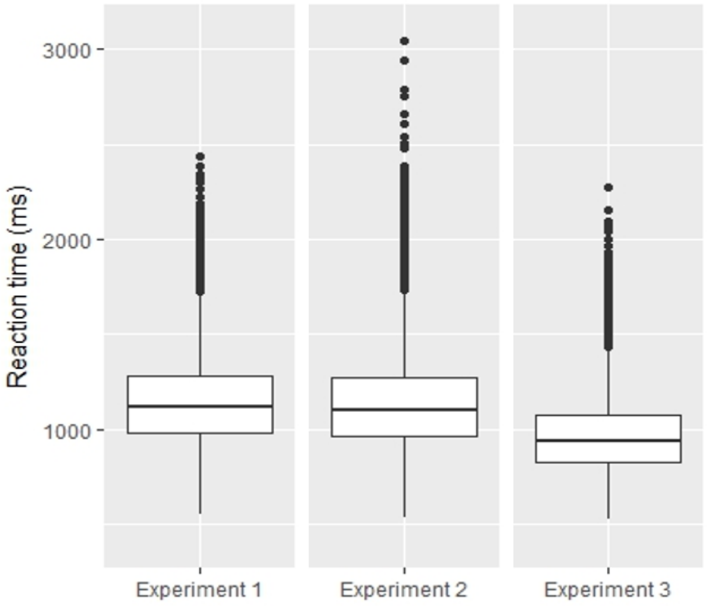


Figure 3.5. Distribution of RTs across experiments: For each box, each horizontal line represents the first and third quartiles, and the median. Data point represented by the whisker lie outside the middle 50%.

Figure 3.5 shows that RTs in Experiment 3 (M=970ms, SD=202.51) are shorter than Experiment 1 (M=1,144ms, SD=231.63) and Experiment 2 (M=1,139ms, SD=253.13), despite the additional color identification task. This is in accordance with multiple participants' responses in the exit survey, who replied that the purpose of including color identification might have been to raise attention during the lexical decision task. This is possible because hearing a color term could block reverberation of the lexical decision item in the previous trial. It is also notable that the percentage of outlier tokens was higher in Experiment 3 (1.71%) than in Experiment 1 (1.15%) and 2 (1.52%). The asymmetry between outlier ratio and RT seems to indicate that the task in Experiment 3 was complicated (yielding more outliers) but attention-grabbing (yielding faster RT).

After the outlier treatment, data from distracter trials were removed, leaving 3,519 tokens from critical trials for RT analysis. In Figure 3.6, RTs from this dataset are split into two subsets and plotted by word age and guise.

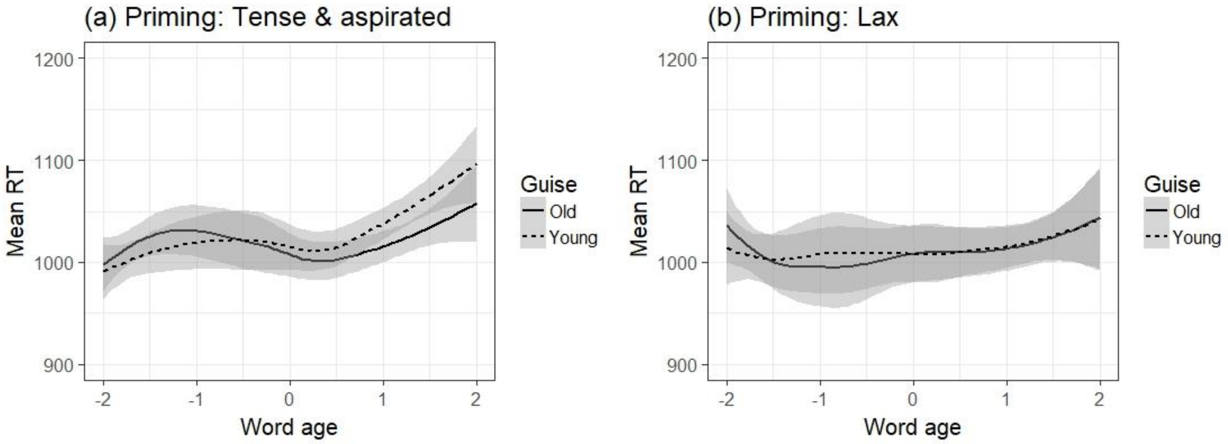


Figure 3.6. Mean RTs by word age and guise: Trials preceded by (a) TENSE or ASP priming (N=2,373) and (b) LAX priming (N=1,146). Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

Figure 3.6 (a) presents data for the critical trials preceded by ASP/TENSE (dataset (a)). There are some notable trends. Recognition of words with word age around -1 was delayed due to an effect of word duration (consistent with Experiment 1 and 2). Thus, a main effect of word age seems to be attenuated by the duration effect; RTs for words with lower word age do not appear to be substantially shorter. In addition, a crossover interaction is evident between word age and guise; responses for words with lower word age were faster when preceded by the younger guise, while words with higher word age were recognized faster when preceded by the older guise, supporting the main hypothesis. However, none of the patterns outlined above are found in the dataset (b) (i.e., trials preceded by LAX presented in Figure 3.6 (b)).

A linear mixed effects model was fit to RTs in the dataset (a). Since the maximal random effects structure was fit in this model without a failure to converge, only results from the frequentist models (not Bayesian) are reported. Fixed effects tested in the model include talker ID, trial order, trial block, test location, computer station, participant sex, participant handedness, participant dialect, item list, and all two-way and three-way interactions between these factors, but none of them reached significance or improved the model's fit. In the final model, guise (deviation coded), word age, and their interaction were included as test variables. Prime ID (TENSE or ASP, deviation coded) and word duration were also included as control variables. Random effects were maximal, including (1) by-participant and by-item random intercepts, (2) by-participant random slopes for guise, word age, and their interaction, and (3) by-item random slopes for guise and an interaction between guise and word age.

Table 3.5. Summary of RT model fit to data from critical trials preceded by TENSE and ASP priming

Model: lmer (RT ~ Guise*WordAge + PrimeID + Duration +
(1+Guise*WordAge | Participant) + (1+Guise:WordAge+Guise | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	743.345	43.148	17.158	<.001
Guise=young	4.340	3.795	1.144	.253
Word age	10.661	9.153	1.165	.244
Prime ID (TENSE vs. ASP)	-3.292	11.442	-0.288	.774
Word duration	0.480	0.065	7.373	<.001
Guise=young : Word age	6.455	2.756	2.342	.019

In Table 3.5, no significant main effects are found for guise (*p*=.253) or word age (*p*=.244). The lack of a main effect of word age seems to be due to a smaller number of the critical items used in Experiment 3 than in Experiment 1 and 2. I interpret the lack of a guise effect as evidence that age cues in a single phonetic variant are not sufficient to induce a main effect of talker age like that shown in Chapter II. Prime ID (TENSE vs. ASP) also failed to reach significance as a main effect (*p*=.774) or in interaction with guise or word age, indicating that the effect of the manipulation was not different, no matter which stop category (aspirated or tense) was used as the prime. The effect of word duration was significant (*p*<.001); participants took longer to respond to longer words. Along with these effects, the results indicate a significant interaction between guise and word age (*p*<.05); when primed by a young-associated phonetic variant, participants took longer to recognize a word with higher word age.

To test whether trials preceded by LAX tokens were influenced by the age-cue manipulation in ASP/TENSE tokens produced by the same talker, a separate model was fit to the dataset (b), with the same fixed effects (excluding Prime ID) and random effects (see Table 3.6).

Table 3.6. Summary of RT model fit to data from critical trials preceded by LAX priming

Model: lmer (RT ~ Guise*WordAge + Duration +
(1+Guise*WordAge|Participant) + (1+Guise:WordAge+Guise|Item))

	Estimate	Std. Error	t value	p value
(Intercept)	777.264	64.341	12.080	<.001
Guise=young	-2.689	5.706	-0.471	.637
Word age	20.837	14.180	1.469	.142
Word duration	0.373	0.094	3.955	<.001
Guise=young : Word age	-0.841	3.875	-0.217	.828

As shown in Table 3.6, the results from this model show no significant interaction between guise and word age (*p*=.828). This lack of an effect does not appear to stem from the smaller number of tokens (N=1,146 in LAX versus 2,373 in TENSE + ASP) since the effect still fails to reach significance even when each response is counted twice (making the number of tokens for the two models comparable) (β=-1.405, s.e.=3.677, *p*=.702). It also does not appear to be due to the smaller number of critical items (N=36 in LAX versus 74 in ASP/TENSE), since the main effect of word age is not smaller in the LAX model (*p*=.142 in Table 3.6, *p*=.092 in the double-counted model) than in the ASP/TENSE model (*p*=.244).

3.5.3. Individual difference in perceived talker age

As a summary of the results in the previous section, a significant interaction between guise and word age is found from the RT analysis of the ASP/TENSE trials, but not in the LAX trials, meaning that the effect of the ASP/TENSE prime is not generalized to the trials preceded by the LAX tokens. One possibility for the lack of the effect in the LAX trials is that listeners may have not tracked the link between realizations of the priming stimuli and each female talker (even though they were explicitly told that there were two female talkers), a problem which will be explored in future research. However, it is also possible that activation of social information is not necessary to observe an effect, as hypothesized in Section 3.3.

Regardless of whether listeners tracked each talker's identity, the lack of an effect for the LAX tokens is interpreted as stemming from either a failure to activate abstract age-based information about the talker, or a failure of the activation to flow from the abstract social information to tokens not immediately preceded by the relevant sociophonetic cues. To test the first of these interpretations further, this section reanalyzes the RT data using perceived talker age obtained from the exit survey (see Section 3.4.3) on the assumption that listeners who perceived greater age difference across talkers at the end of the experiment had been more likely to be sensitive to the relation between the variant and age during the task.

Since perceived age was obtained from participants' ratings on the re-presented priming stimuli, with their attention brought to the question of age, it is not a direct indicator of activation of age information during the task. However, since participants were exposed to only one guise per talker, we can measure the difference in age perceived by each individual across talkers (*perceived difference*, henceforth). Then, using these post-hoc values, we can examine the degree to which age information could be extracted from the phonetic cues and whether the age-congruence effect was enhanced when listeners had greater awareness of the phonetic-social relation.

In addition, if there is substantial variation in perceived difference across participants, it will be worth reexamining the data from those whose age perception was influenced to a greater degree, testing whether the age-congruence effect transferred to LAX trials for these participants. This is because these participants – compared to those with less sensitivity – were more likely to activate social information upon hearing the age-related phonetic cue, and if so, they may have reactivated it upon hearing the identical voice, even though the phonetic cues were not present.

Perceived difference was obtained by subtracting perceived age for the talker in the younger guise from that for the talker in the older guise. Means of perceived age and perceived difference are summarized across the female talkers and guises in Table 3.7.

Table 3.7. Means of perceived age and perceived difference of female talkers by talker's ID and guises

Experimental list	Talker ID	Guise	Perceived age (mean)	Perceived difference (mean)
List 1&4 (N=21)	A	Older	36.33	11.1
	B	Younger	25.24	
List 2&3 (N=19)	A	Younger	24.47	5.42
	B	Older	30.42	
All participants	A or B	Older	33.53	8.4
		Younger	24.88	

As shown in Table 3.7, mean perceived age differed consistently with the manipulation of the guise; participants rated the older guise with a higher value than the younger guise for both talkers, and so mean perceived difference is a positive value for participants exposed to either set of item lists. At the same time, there appears to be variation across talkers; Talker A showed a greater difference than Talker B, indicated by the greater mean perceived difference when Talker A was presented with the older guise (i.e., experimental lists 1 and 4). Additionally, individuals showed substantial variation of perceived difference across participants; with the mean difference of 8.4 years and median of 5, the maximum was 42, whereas the minimum was -5. The standard deviation was 11.153. In a mixed effects model fit to perceived age, a significant main effect of guise was found (β=-8.521, s.e.=1.756, p<.001), indicating that participants' ratings for talker age was lower for the younger guise, compared to the older guise. This model included guise, talker ID, and their interaction as fixed effects, as well as a by-participant intercept[29].

To test whether the age-priming effect found in Section 3.5.2 was conditioned by individuals' difference in perceived age for the two talkers, a linear mixed effects model was fit to RT from the ASP/TENSE trials. Included as test variables were perceived difference (continuous), guise (binary, deviation coded), word age (continuous), and their 3-way interaction. Also included in the model were the same control variables as in the model reported in Table 3.5 (i.e., Prime ID and word duration), as well as the maximal random effects structure. This model indicated a null effect of the 3-way interaction (β=0.036, s.e.=0.257, *p*=.889),

---

[29] Perceived age was also tested as a predictor in a model fit to RT, replacing guise in the model reported in Table 3.5. The random effects included (1) by-participant and by-item random intercepts, (2) a by-participant slope for word age, and (3) a by-item slope for perceived age. In this model, a significant interaction between perceived age and word age was found (β=-0.608, s.e.=0.298, *p*=.041).

suggesting that the interaction between guise and word age was not significantly enhanced by an increase in perceived difference. In other words, listeners' awareness of age-related talker information did not add to the age-priming effect found in the ASP/TENSE trials (Table 3.5).

Next, in order to reexamine the role of listeners' explicit awareness of the age-related associations in the absence of the age-related phonetic cue in proximity, a separate RT model was fit to the data from the LAX trials for a subgroup of participants (N=21) whose perceived difference was greater than the median (i.e., perceived difference $\geq$ 5), with the same predictors as in the model reported in Table 3.6. No interaction was found between guise and word age in this model (β=0.316, s.e.=5.919, p=.957), suggesting that even individuals who were relatively sensitive to the age cues did not generalize talker information when the phonetic cue was not present in proximity.

3.6. Discussion

Analysis of reaction times in Experiment 3 demonstrates that lexical access is facilitated when primed by a phonetic variant that tends to be produced by the age group that the word is associated with. The results are consistent with the findings in Chapter II which demonstrate that recognition of socially-conditioned lexical items is influenced by socio-indexical phonetic realizations. The results from Experiment 3 add to this by providing evidence that the effect is found even when the cue for talker information is restricted to a single phonetic variant, and when it occurs in a word that precedes the target word. In line with experience-based models, this finding suggests that covariance between phonetic realizations, lexical items, and social information is stored in memory, and lexical access is guided by the link between phonetic realizations and lexical items.

Note that most of the prime and target word pairs were neither semantically related, nor ones that co-occur frequently (e.g., *red* + *elementary school*). If they were, experience-based models would predict that the phonetic priming effect was due to a chunk of two related words that are stored in memory together with phonetic details realized by older or younger speakers. I provide two different interpretations of the results below, focusing on the cognitive mechanism through which the effect could have occurred.

First, since the link between phonetic variants and lexical items is formed through experience with socially-conditioned speech of younger and older people, their interplay can be mediated by activation of social information about the talker, including their age. Under this interpretation, upon encountering a socially-conditioned phonetic realization, representations of abstract social information associated with that realization are activated by distributional probabilities, and the activation then spreads over to the subsequent lexical processing. As a result, lexical access is facilitated when the lexical item is associated with the congruent social information. In addition, once social information is activated, it would be indexed to the talker who produced that variant and could then reasonably be generalized to all instances of speech produced by the same individual, unless the activation is cancelled upon encountering counterevidence. In this case, we might expect listeners to exhibit an effect of the ASP/TENSE priming in the trials with LAX priming, as long as they are identified as the same talker, since social information previously activated when listening to that talker might be reactivated. However, this was not observed, even among those participants who exhibit a relatively strong effect of sociophonetic variants on perceived age (Section 3.5.3).

Alternatively, sociophonetic variants may directly index lexical items that are most likely to be used by the speakers who tend to produce those socially-conditioned realizations. In other words, encountering a variant can boost the probability of encountering an indexed lexical item even when social information is deactivated. Since the signal-based socio-indexical cue is restricted to a single variable and is not found in the word itself, the cue may not be prominent and specific enough to activate age-related social information for all participants[30]. Nevertheless, the direct interplay is conceivable, if phonetic forms and lexical items are closely tied in memory. As for the absence of the effect in the LAX trials, since no age-related phonetic cues are found from the LAX prime, there is no link to index particular age-related lexemes unless social information is activated independently. This interpretation is compatible with previous research demonstrating that perception of socially-conditioned speech segments is affected by mere exposure to social concepts, even though listeners hold no explicit beliefs about the talker (e.g., Hay et al; 2006a; Hay & Drager, 2010).

It is important to note that both processes outlined above are generally consistent with the experience-based models that integrate contextual probabilities of social information. In an exemplar-based approach (Johnson, 1997; Pierrehumbert, 2001, 2002), an input of a particular phonetic token can activate similar exemplars stored in memory, lexical representations composed of the distributions of those exemplars, or even representations of any sort of related information including ones represented at an abstract level. During lexical access, multiple, phonetically-rich representations of a single lexeme compete each other for activation, and various sources of information (including information about the talker and the preceding

---

[30] This may seem contradictory to the findings of perception experiments introduced in Section 3.2.3, which show that listeners adapt to talker-appropriate cue-weightings of Korean stops when provided with top-down social information. However, it does not necessarily deny a bidirectional influence between social information and phonetic cues; if the phonetic cues were salient enough, social information would have been activated.

phonetic detail) can contribute to activation rate. Particularly for the second route (i.e., the direct interplay), encountering a phonetic cue in the preceding word would activate the portion in the exemplar distribution that is similar with the preceding phonetic cue, increasing the activation rate of the set of representations of the target lexeme whose distributional properties are consistent with that portion. Notice that since this process depends on direct comparison between the phonetic properties of the preceding signal and the mental representation, it does not necessarily require reference to social information.

The direct interplay between phonetic and lexical information is also compatible with Bayesian models (Norris & McQueen, 2008; Kleinschmidt & Jaeger, 2015), which posit a single abstract representation for a single lexeme. Upon the retrieval of the phonetic cue, listeners would update their generative model adapting to the distributions of phonetic realizations. During this process, prior experience with the covariance between phonetic realizations and lexical items can contribute to the prior probability, increasing the probability of lexical nodes that had often been selected when similar cue distributions were encountered, even if the target word does not contain the same variant. However, what is not as clear as in exemplar models is that these models tend to specify the detailed normative formalizations for the functions of mediating units[31]. For example, Kleinschmidt and Jaeger's (2015)'s mechanism for sociophonetic adaptation relies on changing listeners' beliefs about the talker, making it unclear

---

[31] In connectionist models, on the other hand, specific details of computational processes are often abstracted away via inter-unit connections, if the link is frequently accessed. Since the abstracted process induces highly automatic responses, activation of social information may not be necessary for phonetic cues to robustly prime lexical items that frequently co-occur, in the word itself (Walker & Hay, 2011; Experiments 1 & 2) or in immediately preceding words (Experiment 3).

whether the probability of words can be directly influenced by prior knowledge about the phonetic-lexical covariation or social characteristics of words[32].

The most likely interpretation given the results is that both routes described above contribute to lexical access. Although not all participants revealed explicit awareness of age-related talker information in the exit survey, perceived age ratings were significantly influenced by the presence of the sociophonetic variant, indicating a possibility that, at least for some participants, social information was activated during the task. However, it is also possible that, for those participants, the activation failed to spread to the subsequent lexical processing, and so – regardless of listeners' awareness – the effect was found when age cues were found from the adjacent signal but not when social information was available through the talker representation. Additionally, the priming effect was not conditioned by the degree of listeners' awareness of the relationship between the variant and age.

It is notable that, unlike in Experiment 1 and 2, there was no effect observed for accuracy. A possible interpretation of this result is that there are different types of attention at play across the experiments. Evidence in support of this interpretation comes from work on visual perception that uses the spatial cuing paradigm. Specifically, there is evidence that cognitive and neural processes that influence perceptual accuracy are different from processes influencing RT (Prinzmetal, Park, & McCool, 2005; Van Ede, de Lange, & Maris, 2012). Prinzmetal et al. (2005), for example, demonstrate that when cues for the target location are given reliably (resulting in what they call *voluntary attention*), it changes the perceptual

---

[32] Their model, however, does not entail awareness of talker characteristics at a conscious level. Since the Bayesian framework does not suppose inter-unit activation as connectionist models do, whether activation of social information is necessary may not be a question relevant to these models.

representation of the target stimulus in accordance with the cued location, affecting both RT and accuracy. On the other hand, when the cue is not so strategically advantageous as to change the representation (e.g., there are chance levels of congruence between cue and target locations) but it still draws visual attention (i.e., *involuntary attention*), RT is improved when the cue and target are in the same position while accuracy is not.

Informed by this result from visual perception, the null effect on accuracy in Experiment 3 may be because the cue for talker age was not as informative as the cue used in Experiments 1 and 2 (i.e., a single variable versus talkers from disparate age groups). In other words, the phonetic cues may not have provided sufficient evidence to change some listeners' belief about the talker, so did not influence the mental representation of the talker. However, processing speed was still facilitated when the phonetic variant cue was congruent with the information indexed to the lexical item. This also means that the phonetic variants may have directly indexed lexical items without activation of the talker's social information.

While results from Experiment 3 and the interpretations provided in this chapter illuminate the role of a socially-indexed phonetic variant in subsequent lexical processing, there are some pending questions that need to be addressed in future research. First, it is necessary to replicate the results even when listeners are fully aware of talker identity to see whether – when given the best chance – listeners can and will generalize talker-based information to the LAX tokens. Second, while the priming effect found in Experiment 3 demonstrates the role of listeners' expectations derived by phonetic cues, this work can be extended by testing whether the effect of a single sociophonetic variable is rapid enough to be observed even when the cue is found within the word itself. Such an investigation will further inform the degree to which speech perception is modulated by acoustic match between socially-conditioned signal and representations.

The results from Experiment 3 also have a methodological implication. In psycholinguistic experiments investigating lexical access, potential influences of socially-conditioned phonetic realizations should be considered even when a phonetic variable occurs in the prime word and when the prime is semantically unrelated with the target word. This is particularly the case for targets that are indexed to social characteristics that are related to the phonetic variants.

3.7. Summary of findings

A close link between phonetic variants and lexical items is formed through experience with socially-conditioned phonetic realizations. Listeners’ expectations during lexical access are guided by phonetic realizations of a single socio-indexical variable, and recognition is facilitated when the lexical item and the phonetic variant index similar social information (Section 3.5.2).

While the effect appears to be relevant with talker-specific expectations based on the association between the variant and social information, as indicated by the significant effect of guise on perceived age, we also find evidence for an automatic cognitive process where phonetic variants directly index lexical items and activation of abstract talker information is not necessary. First, the effect is only found when the lexical target is preceded by an age-informative phonetic variant in a single trial, but not when preceded by a non-informative token produced by the same talker (3.5.2). Second, the phonetic priming is not modulated by listeners’ sensitivity to the age-related manipulation (3.5.3). Last, the null effect in accuracy (3.5.1) may indicate that the cue to talker information is too subtle to influence talker representation but is sufficient to facilitate recognition speed when it matches the lexical information.

These findings highlight the role that experience-based probabilistic inference plays in lexical processing. Chapter IV explores a different aspect of the probabilistic inference, this time using a lexical identification task (Experiment 4) in which cues for both the talker and the target lexical item are presented before the auditory stimulus. Experiment 4 will use the visual world paradigm to examine the time course of recognition processes, focusing on listeners' strategic use of prior expectation formed by pre-activated age-related associations.

CHAPTER IV

Time Course of Expectancy-Driven Lexical Recognition

The experiments in the previous chapters demonstrated a lexical processing advantage that arises when the lexical item and its phonetic realization are indexed to congruent age information. These experiments focused on how socio-indexed phonetic detail is utilized during the probabilistic inference of an *unknown* lexical item. However, lexical items (as well as talker information) are often predictable in the real world. Chapter IV examines such a circumstance – in which listeners are presented with two lexical candidates and the talker's voice before hearing the onset of the target word – relating to questions about the role of age-related prior expectancy and the time course of socially-indexed phonetic processing of spoken words. For example, what kinds of expectations are formed by prior exposure to the word that may be socially congruent or incongruent with the talker? Will processing of a signal onset momentarily ambiguous between two cohort candidates be influenced by socially-congruent realizations in the same way as access to an unknown lexical item, even when the lexical items and the talker are highly predictable? If so, can we find real-time evidence for the processing advantage as segmental information unfolds?

4.1. Experiment 4: Eye fixations during identification of age-related words

To explore the questions outlined above, I conducted a lexical identification experiment using the visual world paradigm. In each trial, two orthographic lexical candidates with varying word age and a shared onset syllable were presented on the screen, followed by a frame sentence that cued the listeners to one of the four different talker's identity (two older and two younger talkers) prior to the onset of the target word. Then, one of the two candidate words was produced by that talker, and listeners identified the auditory target by mouse click (see Section 4.2 for more detail about the method). It was hypothesized that listeners would form expectations about the words based on the age-related cues in the frame sentence and that, as a result, they would gaze the lexical items congruent with the voice with higher frequency until the target was disambiguated by the auditory signal (see 4.1.2 for the predictions in detail).

I made the hypotheses outlined above because both exemplar-based and Bayesian-based models predict them. In an exemplar-based account, exposure to the frame sentence pre-activates the candidate word that is most strongly indexed to phonetic cues in the frame sentence, whereas expectancy for a word that is incongruent with the talker information would benefit less from stored exemplars. Likewise, a Bayesian account (e.g., Kleinschmidt and Jaeger's (2015) ideal adaptor model) would predict that, if the prior experience of a lexical candidate is probabilistically consistent with the phonetic properties of the talker, a generative model for a previously encountered talker would be easily generalized to the current talker's model. On the other hand, when it is unsure what types of generative models are adaptable to interpret an age-incongruent candidate, listeners would have to adapt to a novel situation, resulting in delayed recognition. To foreshadow the results, the outcomes predicted by these models were not observed, raising questions about the interplay between pre-activated social information and unfolding sociophonetic cues, and pointing to listeners' ability to tune their expectation according to prior cues.

This section is dedicated to provide information about how the visual world paradigm is used to explore the online processing of spoken words (Section 4.1.1) and to state the hypotheses

(4.1.2). Next, I will provide a description of the method (4.2) and the results (4.3 for RT data and 4.4 for eye fixation data). In Section 4.5, the unexpected results are discussed in light of the experience-based models and methodological issues.

4.1.1. The visual world paradigm

In the lexical decision tasks presented in Chapters II and III, reaction times measured by button press include additional processing time for a metalinguistic decision whether the input was a real word. However, this process is not required in natural speech perception. Due to the difference in the timing between lexical access and button press, it is unclear whether the influence of sociophonetic information on lexical access occurs during a pre-access stage (i.e., concurrently with the mapping of auditory input), or at a post-access stage (i.e., while the representation accessed by phonetic parsing is reevaluated by social information), or both.

As mentioned in Chapter I, the use of an online measurement of processing behavior can provide further insight into the time course of socially-conditioned lexical processing. Experiment 4 adopts an eye-tracking method, also known as *the visual world paradigm* (VWP, henceforth) (Tanenhaus et al., 1995; Allopenna, Magnuson, & Tanenhaus, 1998; Huettig, Rommers, & Meyer, 2011), which has been used in a wide variety of psycholinguistic work. The underlying logic in the VWP is based on a linking hypothesis between eye fixation and the activation level of the entity being fixated (Tanenhaus et al., 2000). That is, the amount of eye fixation on a particular object, or an area of interest (AOI), – typically measured by the probability of fixation at a given time point – is proportional to the amount of listeners' attention paid to the object. Therefore, researchers can test the extent to which listeners' attention is influenced by linguistic materials in the presence of visual stimuli manipulated across experimental conditions.

Among widely used experimental designs in the VWP, one can induce listeners' expectations about utterance-level constraints on either semantic/syntactic properties of words (e.g., Kamide, Altman, & Haywood, 2003; Dahan & Tanenhaus, 2004) or the talker's perspective (e.g., Kamide, et al., 2003; Chambers, Tanenhaus, & Magnuson, 2004; Hanna & Tanenhaus, 2004), by manipulating pragmatic congruence between the frame sentence and the visual world (see also Van Berkum et al., 2008; Tesink et al., 2009 for an ERP method). For example, Kamide et al. (2003) manipulated the verb in the frame sentence (e.g., "The woman will *spread* the butter on the bread." or "The woman will *slide* the butter to the man.") to demonstrate that sentence-level interpretations are immediately predicted by a combination of the verb meaning and visually-provided constraints about the location of the referred entities (i.e., location of the bread and the man).

The VWP studies also have established methods for testing immediate influences of sub-phonemic details of the acoustic input by using either speech produced by multiple talkers (e.g., Creel et al., 2008; Creel & Tumlin, 2011) or within-talker manipulation of phonetic parameters (e.g., Dahan, Tanenhaus, & Chambers, 2002; McMurray et al., 2002; Dahan & Tanenhaus, 2004). In McMurray et al. (2002), for example, listeners heard stop-initial words in isolation, with their onset VOTs manipulated on a 0-40ms continuum along a minimal pair (e.g., *bomb* and *palm*). While their results for the identification responses (measured by mouse clicks on pictured objects) revealed the typical pattern of categorical perception (Liberman et al., 1957), the eye fixation data showed a gradient effect of fine-grained acoustics; fixations to the competitor word (e.g., looks to *palm* in a trial in which *bomb* was clicked on) increased as the VOTs of the stimuli approached to the boundary between the two categories.

Another popular technique in the VWP that is also used for Experiment 4 is to present members of a cohort set (Marslen-Wilson, 1987; 1989) to induce signal-based lexical competition during a lexical identification task (e.g., Allopenna et al., 1998; Huettig and McQueen, 2007). A classic example is Allopenna et al. (1998), as discussed in Section 1.1.2, who demonstrated that acoustic similarity in word onsets, as well as rhyme parts, is simultaneously integrated during lexical access, influencing activation rates of lexical competitors.

Experiment 4 was designed combining these three VWP methods to examine the time course of an age-congruence effect and test the predictions of experience-based models. I expected that listeners' eye fixations during lexical competition amid the initial phonetic ambiguity should be influenced by age associations between lexical candidates and the talker information carried in the frame sentence.

4.1.2. Word age conditions and predictions

It was assumed that anticipatory eye fixations during the phonetic ambiguity would be conditioned by the high predictability for the target lexeme and the talker. Specifically, when the talker's voice is provided, listeners would form more explicit expectancy for age-congruent combinations than incongruent combinations. In addition, as soon as the auditory input begins, age-congruent phonetic realizations would add a boost, guiding access to an age-congruent candidate. Four conditions of word age contrasts were created as outlined in Table 4.1 to test two different hypotheses derived from the same prediction that lexical processing would be affected by age associations between words and talkers concurrently with the unfolding signal.

As a main hypothesis, I predicted that fixation probability for the target would be higher before the auditory stimulus is disambiguated (*early region*, henceforth), when the word is produced by an age-congruent talker than when it is produced by an incongruent talker. Since the listeners would eventually identify the target immediately after perceiving the disambiguating cue, the effect is expected to fade during the time course between the disambiguation point and the mouse click (*late region*). Two types of *target-contrast conditions* (the first two lines in Table 4.1) are designed to test the main hypothesis. In these conditions, auditory stimuli (i.e., targets) are composed of young words (i.e., young target condition) or old words (i.e., old target condition). Cohort competitors in these items are ones that do not belong to the word age group of the target. For example, an old word target is paired with a neutral or young word competitor and vice versa for a young target[33].

Table 4.1. Summary of word age conditions and hypotheses

Contrast made for	Condition Name	Word age		Prediction for fixation probability before phonetic disambiguation
		Target	Competitor	
Target	Young target	Young	Neutral-old	Higher on target when talker=younger
	Old target	Old	Neutral-young	Higher on target when talker=older
Competitor	Young comp	Neutral	Young	Higher on comp when talker=younger
	Old comp	Neutral	Old	Higher on comp when talker=older

---

[33] Due to a limited number of age-related words and words that had been recorded for experiments reported in Chapter II, it was impossible to make an ideal contrast of word ages in all conditions. For example, the target-contrast conditions contain both age-related and neutral competitors, and the competitor-contrast conditions could not use infrequent words as targets. Nontheless, I decided to maintain the homogeneity of the auditory stimuli with Experiments 1 and 2 for the sake of comparability in the interpretations of results.

A secondary hypothesis was made based on an assumption that the age-congruence effect would arise even when the candidate attended to due to age-congruence is not produced as the target. I predicted that, when talker age matches the word age of a competitor in the visual scene, the competitor would draw attention before the disambiguation point, delaying target fixation. As a result, there would be a higher frequency of looks to the competitor that matches the talker's age, and a lower frequency of looks to the target. Accordingly, items in the *competitor-contrast conditions* (i.e., young competitor condition and old competitor condition) have competitors of varying word age, while the auditory targets are composed of age-neutral words that occur with high frequencies in the Sejong Corpus. However, the prediction for these conditions was less clear because listeners would also have sufficient prior experience with the highly-frequent age-neutral target words produced by speakers from either age group. I expected that participants could form expectancy about phonetic realization of those words equally well regardless of talker age[34]. Also note that the presence of the competitor-contrast conditions neutralizes the probability of an age-related word to be produced as an auditory target, so that items in these conditions play as filler items for the target-contrast conditions, and vice versa.

Predictions for mouse click RTs are in line with the hypotheses made for eye fixations. I predicted that responses would be facilitated for an age-related target when it is produced by an age-congruent talker. Conversely, responses to an age-neutral target might be delayed when the

---

[34] Although it was not explicitly stated in Chapter II, participants in Experiments 1 and 2 responded to the frequent, age-neutral words (i.e., items categorized as fillers in those experiments) faster and more accurately than the age-related items, regardless of talker age. In addition, Dahan et al.'s (2001a) VWP study demonstrated that high-frequency words draw more frequent fixations even when they are not used as the target word but as a cohort competitor or an unrelated distracter. This was not a problem, however, in the target-contrast conditions because neutral competitors in those conditions were not frequent words.

word is produced by a talker whose age is congruent with an age-related competitor in the visual scene.

4.2. Method

4.2.1. Lexical stimuli

96 bisyllabic words chosen from Chapter II were used as lexical targets. The auditory stimuli were produced by the four talkers in Experiments 1 and 2 (OM, OF, YM, YF), and composed of 24 young, 24 old, and 48 age-neutral words, based on the categorical word age distinction in 2.1.1.2. Each lexical target was paired with another word to make up 96 paired items to appear as visual stimuli. Among them, 48 were critical items (12 items in each condition), for which the visual scene presented a target and a competitor, a word that is disambiguated at either the onset consonant or the nucleus vowel of the second syllable. 48 filler items were added to distract listeners' attention to the phonological competition at the first syllable. The competitor in the filler items was a word that is disambiguated from the word onset. Word ages of the filler items were also manipulated in a similar way. 24 items contained an age-associated target (12 old words and 12 young words) and a frequent neutral competitor. The other 24 items contained a frequent neutral target and an age-associated competitor (12 old and 12 young words). See Appendix II for all lexical stimuli.

Some of the competitors were ones that had not been included in Chapter II. To ensure that word ages of all lexical stimuli are rated by the same population, word ages were obtained during an exit survey using the stereotype-based word age rating method in 2.1.1.1. All word age statistics used below are based on them. Table 4.2 provides a summary of the word age distribution of the targets and competitors in each condition of the critical items.

Table 4.2. Summary of word age distribution (min., mean, and max.): Young words refer to items with a word age from -2 to -1, neutral words are from -1 to 1, and old words are from 1 to 2. The values in the 'difference' column represent the mean word age difference between the targets and competitors, calculated by subtracting word age of the competitor from that of the target per item and then calculating the mean across the items.

Condition	Target	Competitor	Difference
Young target	Young words (N=12) -1.97, -1.81, -1.51	Neutral (N=9), Old (N=3) 0.14, 0.77, 1.57	-2.57
Old target	Old words (N=12) 1.17, 1.45, 1.97	Neutral (N=7), Young (N=5) -1.84, -0.78, 0.49	2.24
Young comp	Neutral words (N=12) -0.38, -0.11, 0.00	Young (N=10), Neutral (N=2) -1.97, -1.55, -0.63	1.44
Old comp	Neutral words (N=12) -0.54, -0.06, 0.24	Old (N=11), Neutral (N=1) 0.95, 1.43, 1.84	-1.49

The duration of the ambiguous region for each item was measured manually using both auditory and acoustic analysis of waveforms and spectrograms. In case acoustic cues detected by spectral visualization differed from my own auditory perception, auditory cues were prioritized. Coarticulatory cues were taken into account with caution. As reviewed in Chapter I, listeners can use coarticulatory information to predict the word concurrently, such as pre-nasalized vowels (Lahiri & Marslen-Wilson, 1991; Beddor et al., 2013). Thus, for example, for an item where the target [ʌmʌm] – a kinship term to refer to one's 'mom' in a respectful manner – is paired with the competitor [ʌpʰɯl] – a shorthand form for the English word '*application*' – the disambiguation point was set to the point at which nasalization was audible in the onset vowel [ʌ] in [ʌmʌm]. Conversely, there were items for which the two candidates shared the same phonetic feature at the onset of the disambiguating segment (e.g., [k*ultɕɛm], a coined word for ‘fun’, paired with [k*ultɕ^{h}a], ‘honey tea’). In these cases, the disambiguation point included the time point at which the shared feature is observed (i.e., the release point of the affricate /tɕ/).

There was variation in the lengths of the ambiguous region across items (min.=145ms, mean=306ms, max.=464ms, s.d.=75.539). The mean durations of ambiguous regions are summarized by talkers and word age conditions in Table 4.3.

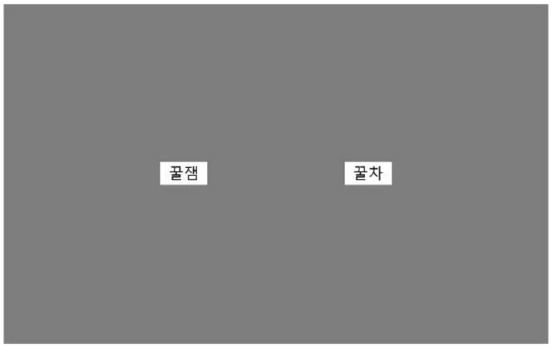
Table 4.3. Mean duration of ambiguous regions (in ms) summarized by conditions and talkers

Condition	OM	OF	YM	YF	Grand mean
Young target	341	371	336	365	353
Old target	261	287	286	306	285
Young comp	306	336	294	330	316
Old comp	248	275	266	287	269
Grand mean	289	317	296	322	306

Recall that there was a non-significant effect of an interaction between word age and talker age on durations of word stimuli used in Experiments 1 and 2 (*p*=.075) (see Section 2.3.1). In order to test whether the durations of the ambiguous regions are influenced by the age associations, a mixed effects model was fit to the durations of the auditory stimuli used in Experiment 4. Included as fixed effects were main effects and an interaction of talker age (binary, deviation coded) and target word age (continuous), along with talker gender (binary, deviation coded) as a control variable. A by-item random intercept and slopes for target word age and its interaction with talker age were also included. The output of the model indicated a non-significant effect of an interaction between talker age and target word age (β=.-1.294, s.e.=7.183, *p*=.086).

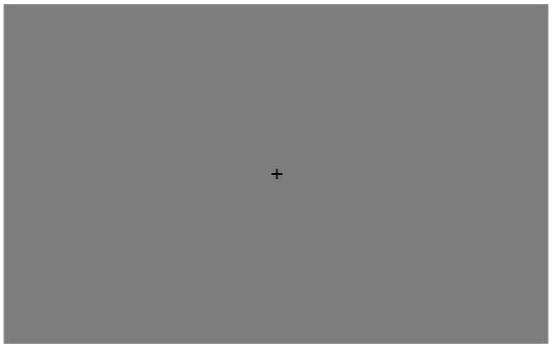
4.2.2. Procedure

Participants were seated on a height-adjustable stool in front of a computer monitor screen. An eye-tracking device (see Section 4.2.4) was attached at the bottom of the monitor. Participants read instructions on the monitor and familiarized themselves with the task through eight practice items. Next, a calibration procedure was performed, during which participants were directed to fix their posture at an appropriate distance from the monitor (60-70 cm) and look at nine focal points on the screen successively. Deviations from the focal points were automatically recorded. The calibration process was repeated for each participant (up to five times per participant) until fixation was captured within a deviation range of 1 degree of visual angle, and then the word identification experiment began. Each trial was composed of three slides presented on the monitor, as shown in Figure 4.1.



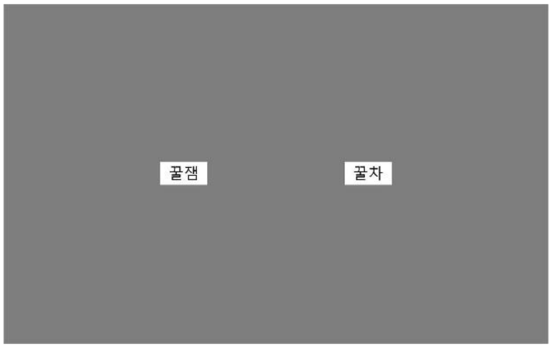
(a) Preview

- Slide duration: 3,000ms
- Audio: None
- **Lexical candidates shown**



(b) Fixation cross

- Slide duration: 1,500ms
- Auditory prime: “The word is…”
- Audio length: 888ms for all the four talkers
- **Talker information provided**



(c) Test slide

- Slide duration: 3,000ms
- Auditory stimulus: A target word with various word age and length
- **Lexical target provided**

Figure 4.1. Example visual stimuli of word identification task

Each trial began with a 3,000ms preview of the visual scene (Figure 4.1 (a)), which presented the lexical candidates. Participants were instructed to look at the words and think what each word means during the preview scene. This was done so that the lexical information was fully activated with before hearing the target word (see below in this section for more detail). Each word in the visual scene was written in black on a white rectangle (145 pixels * 75 pixels) against a grey background, and the two boxes were placed far apart enough to decrease the possibility that the two words would be recognizable at a glance. AOIs of the target and competitor were set to a larger rectangular region (340 pixels * 200 pixels) proportionally encompassing each side of the two rectangles.

The preview slide was followed by a 1,500ms presentation of a fixation cross located at the center of the screen (Figure 4.1 (b)). Along with the fixation cross, a frame sentence, *ipen tanenun*…, “The word is…” was played through headphones 200ms after the fixation cross appeared, providing information about the talker in the given trial. Participants were instructed to look at the cross while it was on the screen to prevent pre-fixations on the AOIs before the auditory input. The frame sentence for each talker was selected from raw recordings obtained in Section 2.1.1.4. The prosodic pattern of the frame sentences was controlled across talkers as much as possible. The duration of the frame sentence was 888ms for each talker.

Finally, the auditory stimulus of the lexical target was provided in the test slide (Figure 4.1 (c)). The visual stimulus was identical to the preview slide of each trial. An auditory stimulus (target word) was played simultaneously with the appearance of the visual stimulus, which was 1,300ms after the frame sentence onset. Participants were instructed to identify the word they heard by mouse-clicking the word on the screen as quickly as possible. The time point and the pixel at which the mouse click occurred was saved automatically. The target slide remained for 3,000ms and automatically changed to the next trial. The position of the mouse cursor was automatically set to the center of the screen at the onset of the first (preview) and the second slide (fixation cross), but not the test slide due to technical limitation.

The final design described above was determined after piloting a different version of the experiment with 36 participants, the method of which differed in various aspects including lexical items, the number of lexical candidates (two or four), size and distance of AOIs, and whether to present the frame sentence while a preview of the lexical stimuli are presented on the screen. There are three important decisions I made based on the pilot data include the following. First, the final set of lexical items were selected in order to maximize the durations of ambiguous regions in which age-associated anticipatory looks would be observed. Second, only two lexical candidates were used to make sure that participants were fully activated with the lexemes and to decrease unintended fixations in search of the target during the early time point. Third, a fixation cross was inserted to prevent pre-fixations, resetting fixations drawn by expectations during the frame sentence[35].

There are some additional decisions made due to using age-indexed words as stimuli. First, written words were used (rather than pictured objects) because most of the lexical items

---

[35] I will discuss the influence of the insertion of the fixation cross on fixation patterns in 4.5.2.1.

were undepictable (e.g., *kkwulcaym*, a slang used to mean ‘fun’). Orthographic stimuli have been tested in previous VWP studies that investigate languages using phonograms (e.g., Kim, Mitterer, & Cho, 2017 for Korean, D'Onofrio, 2015; Koops et al., 2008 for English, and Huettig & McQueen, 2007; Brouwer, Mitterer, & Huettig, 2012 for Dutch) or pictograms (e.g., Wiener & Ito, 2015 for Chinese). Second, the preview slide was inserted to assure that the lexical properties (including word age) are fully activated. There is evidence that orthographic stimuli in the VWP drive listeners' attention mainly to phonological processing, with semantic properties not activated as much as in tasks using pictured objects (Huettig & McQueen, 2007)[36]. Last, the talker information was provided in a frame sentence (rather than in a word in isolation) in order to test age-related anticipatory fixations as a function of expectations prior to the target onset (in line with Experiments 1 and 3), rather than immediate integration of the age-indexed phonetic realizations (in line with Experiment 2). This was because I was uncertain about how rapid the age-congruence effect would occur at the time of designing Experiment 4. However, since a rapid influence with no resort to prior expectations was demonstrated in Experiment 2 using the lexical decision paradigm, it is a question for a follow-up experiment to test whether the integration indeed takes place online or at a post-perceptual phase.

4.2.3. Design

Items were counterbalanced by the position of the target word (left or right) and talker age (older or younger) across 4 lists (2*2) in a Latin Square design. Participants were randomly assigned to one of the four lists, and the stimuli appeared in a random order per participant. Items were

---

[36] Note that it is possible for a listener to click on the target word, even without noticing what it means, but impossible to do so without perceiving the segments.

counterbalanced for talker age but not gender; one group of participants, for example, heard an item in the OM’s voice, while the other group heard that item in the YM’s voice, and vice versa for the OF and YF. The target and competitor were not counterbalanced within a given lexical pair, either.

4.2.4. Apparatus

Data were collected at the LAE eye-tracking Lab, University of Hawaiʻi at Mānoa, using SMI RED 250 eye tracker with sampling rate at 250Hz. The experiment was implemented and run on the SMI Experiment Center software (ver. 3.7.76). AOI settings and data exportation were done in the SMI BeGaze software.

4.2.5. Participants

Data were collected from 29 college students who had lived in Hawaiʻi less than six months at the time of participation. All but four participants listed the Seoul Dialect as their most frequently used dialect. These four participants listed Kyeongsang (N=3) and Jeolla (N=1) as their primary dialect but they also reported being fluent in the Seoul Dialect.

For the mouse click analysis, two participants were removed due to slow reactions (see Section 4.3.1), leaving 27 participants (24 females and 3 males, age: 19-23). For eye fixation analysis, two different participants were removed due to insufficient fixation data (see 4.4.1). Data from 27 participants (25 females and 2 males, age: 19-23) were analyzed in eye fixation analysis.

4.3. Mouse click RT analysis

4.3.1. Results

2,784 mouse click RT tokens were collected from the 96 items by 29 participants. RTs were measured from the time when the lexical target is disambiguated in each trial[37]. Data from two participants (192 data points) were removed who had excessively long RTs in general (means: 1,084ms, 1,148ms). The average RT of all participants was 878ms. Additionally, there was 1 token where the participant failed to make a click during the 3,000ms test slide, 1 token with an incorrect mouse click (click on the competitor word), and 2 tokens with RT over 2,500ms. After removing these tokens, responses that fall out of three standard deviations from the mean by participant were removed (N=14). Finally, data from filler trials were removed, and 1,288 data points were left to be analyzed below.

Figure 4.2 presents the mouse click RT results. Data from the target-contrast conditions (N=647) are plotted by target word age and talker age in (a), and data from the competitor-contrast conditions (N=641) are plotted by competitor word age and talker age in (b).

---

[37] For filler trials, RTs were measured from word onset.

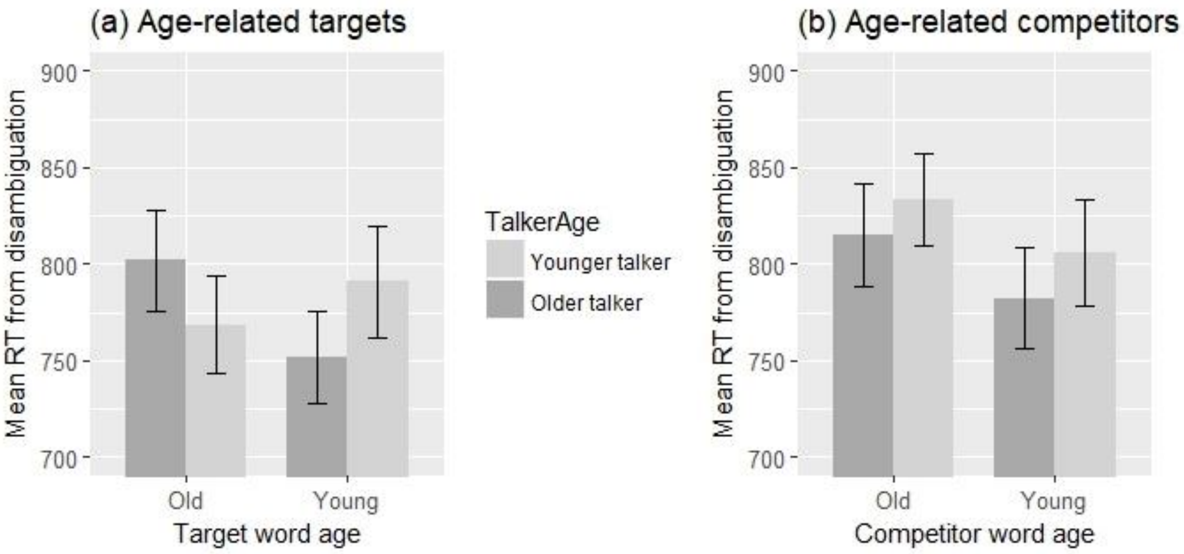


Figure 4.2. Mouse click RT predicted by word age and talker age: data from (a) target-contrast conditions and (b) competitor-contrast conditions

Figure 4.2 (a) shows that when an old word was produced as the target, mean RT was shorter (i.e., the target word was identified faster) when the word was produced by a younger talker than an older talker. In contrast, when a young word was produced as the target, mouse clicks tended to occur faster when the word was produced by an older talker than a younger talker. Surprisingly, listeners identified the target faster when talker age and target word age did not match, an opposite pattern than what was observed in the previous three experiments.

As for the competitor-contrast conditions, no interaction between talker age and word age is observed in Figure 4.2 (b). Since these conditions differed from the target-contrast conditions in that age-related words were present in the scene but not produced as target, it appears that mouse click responses were facilitated by age mismatch only when the age-related word was the word that was encountered.

To test the trend in the target-contrast conditions statistically, a linear mixed effects model was fit to raw RTs using the lme4 package[38]. The model structure was determined using the same method in the previous chapters (see 2.1.2.1). Fixed effects in the final model included talker age, target word age (both binary, centered) and their interaction. Other fixed effects tested prior to constructing the final model include target location, trial order, experimental list, word duration, duration of the ambiguous region, and word age difference between the target and competitor (instead of target word age). All tested variables were centered. Random effects included by-participant and by-item intercepts, and by-item slope for the interaction between talker age and word age.

Table 4.4. Summary of model fit to mouse click RT (data: young and old target conditions)

Model: lmer (AlignedRT ~ TalkerAge*TargetAge +
(1 | Participant) + (1+TalkerAge:TargetAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	778.639	17.638	44.147	0
Talker age=young	3.006	21.490	0.140	.889
Target word age=young	-14.684	18.644	-0.788	.431
Talker age=young : Target age=young	**74.535**	**42.980**	**1.734**	**.083**

---

[38] For both mouse click RT and eye fixation analyses, only results from frequentist models are reported (i.e., no Bayesian models were fit to include the maximal random effects structure). This is because fitting a model with a larger random effects structure can lose statistical power in exchange for decreasing the Type I error rate, and this is especially true when the sample size is small (Matuschek, et al., 2017). Experiment 4 had a relatively small number of observations for each possible combination of word age and talker age (6 per participant).

As shown in Table 4.4, main effects of talker age (*p*=.889) and target word age (*p*=.431) were not significant. An interaction between them shows that participants tended to respond more slowly when talker age matched word age, but the effect was not significant (*p*=.083).

A separate model was fit to the data from the competitor-contrast conditions. Fixed effects included talker age and competitor word age (as main effects and in interaction), and random effects were identical with the model in Table 4.5.

Table 4.5. Summary of model fit to mouse click RT (data: young and old competitor conditions)

Model: lmer (AlignedRT ~ TalkerAge*CompAge +
(1 | Participant) + (1+TalkerAge:CompAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	809.354	16.691	48.489	0
Talker age=young	18.860	17.214	1.096	.273
Comp word age=young	-15.661	11.834	-1.323	.186
Talker age=young : Comp age=young	**4.989**	**17.214**	**0.290**	**.772**

As shown in Table 4.5, along with insignificant main effects of talker age (*p*=.273) and word age (*p*=.186), talker age also did not interact with word age of the competitor (*p*=.772). Thus, word age of the competitor appears to have little influence on RT, both as a main effect and in interaction with talker age.

In the target-contrast conditions, responses tended to be facilitated by age mismatch but the effect did not reach significance. However, since the competitor-contrast conditions contain age-neutral targets, it is also possible to collapse the data from all conditions and test the influence of target word age with a greater statistical power, by treating word age as a continuous variable and including more data points. Although the target-contrast conditions and competitor conditions differ in how word ages were manipulated between lexical candidates, such an analysis is valid because an effect of target word age on RT would not be substantially distorted by competitor word age, given the null effect in the competitor-contrast conditions.

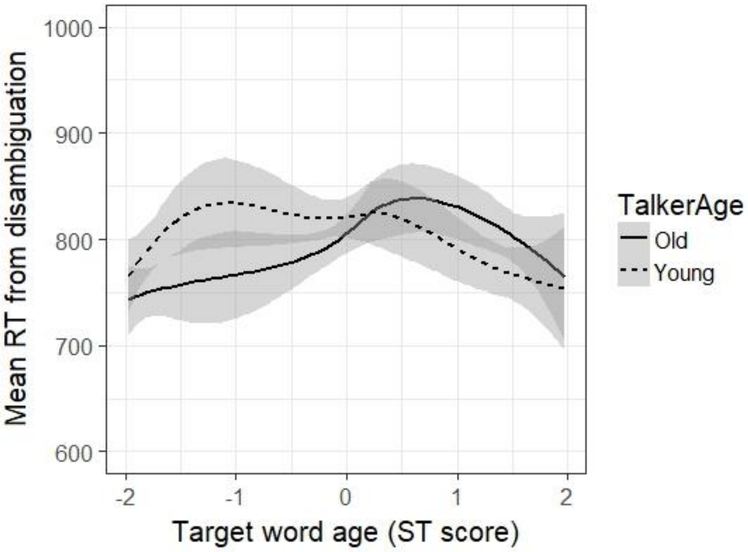


Figure 4.3. Mean RT predicted by continuous word age and talker age (data: all conditions)

In Figure 4.3, RTs were generally shorter for target words that are strongly associated with age (i.e., words with word ages close to -2 or 2) than words that are slightly associated with age (i.e., words with word ages close to -1 or 1). Since lexical candidates were pre-activated, I interpret that listeners paid greater attention to socially-salient items throughout the task, resulting in faster identification responses. In addition, responses to age-neutral targets (i.e., words with word ages close to 0) were also faster than the slightly age-related items but slower than the strongly age-related items. Since these items were the frequent words used in the competitor-contrast conditions (see Section 4.1.2), this seems to indicate that identification occurred faster for words with higher frequency.

Along with these trends, a cross-over interaction between talker age and target word age is found; listeners identified the target faster when talker age and word age were incongruent. This trend was tested in a model fit to RT. Fixed effects included talker age, target word age (both centered, as main effects and in interaction), and trial order (continuous). Random effects were maximal, including (1) by-participant and by-item intercepts, (2) by-participant slopes for talker age, target word age, and their interaction, and (3) a by-item slope for talker age.

Table 4.6. Summary of model fit to mouse click RT (data: all conditions)

Model: lmer (AlignedRT ~ TalkerAge*TargetAge + TrialOrder +
(1+ TalkerAge*TargetAge | Participant) + (1+TalkerAge:TargetAge+TalkerAge | Item))

	Estimate	Std. Error	t value	p value
(Intercept)	810.796	17.465	46.424	0
Talker age=young	13.159	13.051	1.008	.313
Target word age (cont.)	4.465	6.547	0.682	.495
Trial order	-0.118	0.048	-2.430	.015
Talker age=young : Target age (cont.)	**-32.581**	**12.308**	**-2.647**	**.008**

As shown in Table 4.6, a significant main effect of trial order (*p*<.05) indicates that click responses were facilitated as the experiment session proceeded. An interaction between talker age and target word age is also significant (*p*<.01), indicating that identification of the auditory target was facilitated when a younger talker (compared to an older talker) produced words with higher word age (i.e., when word age and talker age were incongruent).

In a separate model that included competitor word age (instead of target word age) as a main effect and in interaction, with the same random effects, no interaction between talker age and word age was found (β=6.409, s.e.= 10.745, *p*=.551).

4.3.2. Interpretation

In the previous section, I demonstrated that lexical identification speed measured by mouse click RT was facilitated by age mismatch between the talker and the identified word. The results seem to demonstrate that the identification process was affected by preparatory attention drawn to a word that was unlikely to be produced by the talker. The experiment design meant that socially congruent and incongruent candidates were equally likely to be the target. To be equally prepared for both, listeners' attention seems to have been drawn to the incongruent candidate (i.e., the one which had not been frequently encountered with phonetic realizations similar to the given talker's voice). Once a candidate was attended to, responses occurred more quickly when the attended item was the target even if the phonetic realization was incongruent with prior experience of the word. Such interpretation will be evaluated by eye fixation patterns during the time course of the identification task (see Section 4.5.1 for further discussion of the effect).

4.4. Eye fixation analysis

Prior to running the experiment, I predicted that eye fixations would be drawn to the age-congruent candidate with a higher probability than the incongruent candidate (whether the word eventually turns out to be the target or not). However, the mouse click RT data suggests that attention may be drawn instead to the incongruent candidate. Therefore, this section examines the eye gaze data, focusing on whether the mismatch-driven effect is also observed in fixation patterns.

4.4.1. Data treatment and modeling.

Eye gaze data in the test slide were exported by the SMI BeGaze software and binned into 20-ms samples. All data points in which a fixation occurred were transformed into a binary variable for each of the three AOIs (target, competitor, grey space), indicating whether the participant was fixating the AOI at a given time point. Only fixation data were subject to analysis but data from time points that the tracker did not treat as a fixation (e.g., saccades, blinks, tracking failure) were also exported. Since it is possible that such non-fixation events occurred during an actual fixation event (e.g., a blink while fixating on an AOI), non-fixation events shorter than 100ms (i.e., up to 80ms, which means four consecutive non-fixation data points) between fixations to the same AOI were treated as fixations to that AOI.

Proportion of non-fixation data was calculated for each participant, and data from two participants whose rate of non-fixation data were over mean+2SD were removed for having insufficient fixation data. This left data from 1,296 trials obtained from 48 critical trials by 27 participants. Then, I removed 153 trials (11.81%) in which (1) non-fixation rates were greater than mean+2SD of all trials (N=59), (2) an incorrect mouse click occurred (N=1), (3) fixations were made only on the grey space (N=36), (4) participants were fixating on the target or competitor at the trial onset instead of the fixation cross (N=42), or (5) participants indicated in the exit survey that they did not know one of the lexical candidates (N=15). The process outlined above left 151,302 fixation data points from 1,143 trials for analysis.

The results will be presented in two subsections. First, results from the target-contrast conditions are analyzed, the purpose of which is to test whether fixation to age-related targets are influenced by congruency between talker age and target word age (Section 4.4.2). Second, results from the competitor-contrast conditions are provided and fixations to age-related competitors will be analyzed (4.4.3). In each section, fixation data are analyzed separately for two different time regions. First, data from the early region (from trial onset to disambiguation+200ms[39]) are analyzed to examine anticipatory fixation patterns at the very initial stage; fixation behaviors under phonetic ambiguity would be influenced either by expectations formed prior to the signal onset or by age-indexed realizations prior to the phonetic disambiguation. Second, analysis for data from the late region (from disambiguation+200ms to mouse click) are expected to provide further information about the preparatory effect. If the mouse click results were in fact due to preparatory attention drawn to age-incongruent words, fixation rate for those words would be higher than age-congruent words in the later stage – and this effect would be greater in the late region than in the early region – because the initial mismatch-driven attention would be boosted when programing the response upon hearing a socially-incongruent phonetic realization.

For the statistical analysis, binomial mixed effects models[40] were fit to a binary variable of AOI fixation (whether the AOI was fixated at a given time bin). Using the method described in 2.1.2.1, fixed effects and random effects best supported by the data were selected for each model separately. Word duration, ambiguous region duration, and fixation state[41] were tested as fixed effects in each model but they did not reach significance or improve the model's fit. Test

---

[39] In eye-tracking studies (e.g., Allopenna et al., 1998), it is generally assumed that changes in fixation proportions lag behind the auditory input by about 200ms (Magnuson et al., 2013). The fixation delay is in line with the observation that programing and initiating a saccadic movement takes about 200ms (Martin, Shao, & Boff, 1993).

[40] Also tested was empirical logit of looks to target versus competitor, with data aggregated by subject and item, but the model's fit was better when logistic regression model was fit.

[41] Fixation state refers to a binary variable indicating whether target was being fixated in the previous time bin. In Kim et al. (2015), this variable is found to be the best predictor in logistic regression modeling for eye fixation behavior. However, it did not reach significance in my data possibly due to difference in treatment of time; their data were sampled every 4ms, but since my data were aggregated over 20-ms time bins, fixation state was not powerful enough.

variables included in the final models were talker age and word age (both binary[42]) as main effects and in interaction. Time (aligned to the disambiguation point)[43] and target location (left or right) were also included as fixed effects (control variables). Additionally, an interaction between time and the test variables was included only when it was both significant and improved the model's fit. The three binary predictors (talker age, word age, and target location) were centered, and time (in milliseconds) was rescaled to seconds to avoid problems caused by different numeric scales used across variables.

4.4.2. Fixation to age-related target.

I begin the fixation analysis by presenting the results from the target-contrast conditions. The proportions of fixations to the target and competitor are presented on a time course of the test slide in Figure 4.4. Since talker age and word age are both treated as binary, conditions are represented by *match* or *mismatch* in Figure 4.4 (see Sections 4.4.2.1 and 4.4.2.2 for results from the young target and old target conditions separately). The match condition refers to the trials in which either a young or old target word was spoken by one of the two congruent-age speakers

---

[42] Both in the target-contrast conditions and in the competitor conditions, word age contrast is made by young words and old words (i.e., young target versus old target, or young competitor versus old competitor). Because neutral words are not included, treating word age as continuous would decrease the statistical power.

[43] The data presented below appear to bear non-linear changes of fixation rates over time in the late region. The polynomial influence of time can be better captured in Growth curve models (Mirman, 2014; see also Wang, Wang, & Malins (2017) for an application for the visual world paradigm), but time was treated linearly for all models in this dissertation, keeping statistical methods consistent across the time regions.

(male and female), and the mismatch condition refers to the trials in which talker age and target word age mismatched.

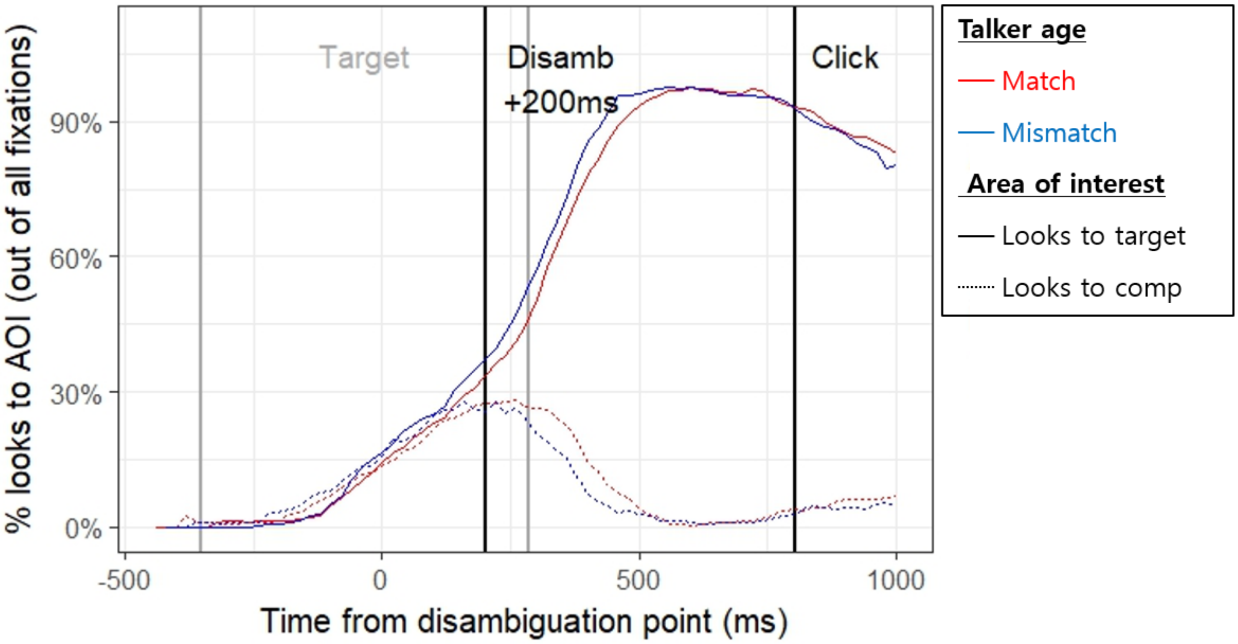


Figure 4.4. Proportions of looks over time (collapsing the young and old target conditions): Talker age is differentiated by color (age-congruent=red, incongruent=blue), and AOI by line type (target=solid, comp=dotted). Time (x-axis) is aligned to disambiguation point (Time=0). Means of target words' onset and offset are indicated by vertical grey lines, respectively. Two black vertical lines (in bold) indicate disambiguation point+200ms and the mean click time of the target-contrast conditions.

As shown in Figure 4.4, mean fixation proportions to target (solid lines) and competitor (dotted lines) began to increase from the trial onset and fixations to the two AOIs diverged approximately at the disambiguation+200ms point (marked by a black vertical line). This is in line with the standard assumption in the visual word paradigm that it takes about 200ms for an eye movement to occur following the perception of auditory signal, and indicates that the listeners began fixating on the target more frequently than the competitor as soon as they heard the disambiguating phonetic cue. The proportion of looks to target remained over 90% for a few hundred milliseconds and began to drop near the mean mouse click point.

Apparent in the region before the mean click time is that target fixation probability was generally higher when talker age mismatched the target word age (blue solid) than when they matched (red solid). The effect appears to be larger in the late region (after listeners perceived the disambiguating cue) than in the early region, indicating that words produced in the incongruent voice began to receive greater attention during the period of phonetic ambiguity, and then the attention further increased during the time region where listeners interpreted the disambiguated signal as a word that mismatched the age-related phonetic cues. Toward the end of the late region, the difference in fixation rates between the matched and mismatched conditions gradually diminished until the click. The non-existence of difference immediately prior to the click is unsurprising since the target eventually had to be fixated to make the click, regardless of congruency.

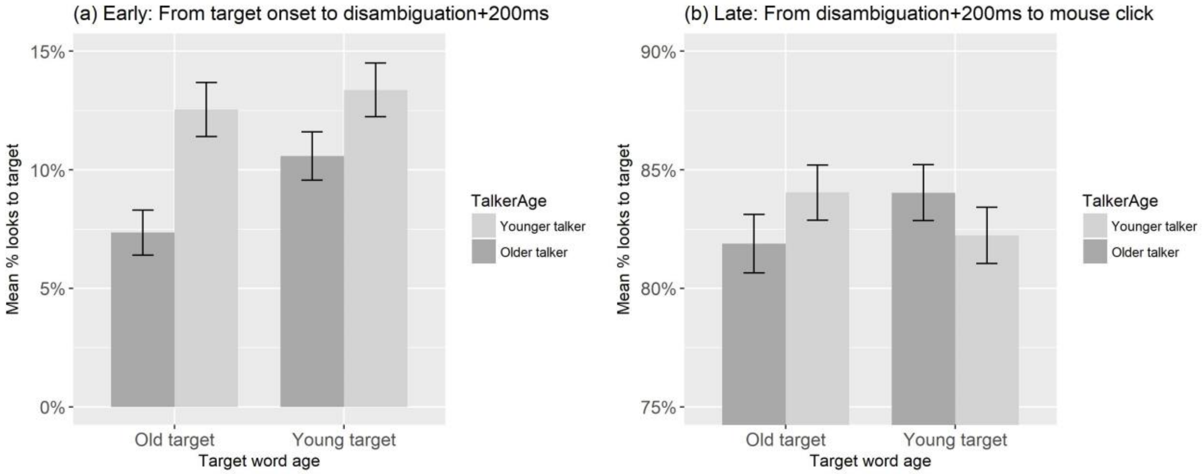


Figure 4.5. Mean target fixation probabilities by word age and talker age: (a) early region, (b) late region. Error bars represent 95% confidence interval of means by participants.

To further examine the difference between the early and late regions, the mean probability of looks to target in the two regions is plotted by target word age and talker age in Figure 4.5. In the early region (Figure 4.5 (a)), it appears that listeners fixated the age-related target more frequently when the talker was a younger talker than an older talker, regardless of word age. However, in the late region (Figure 4.5 (b)), an interaction between target word age and talker age is observed; targets were fixated with a greater probability when the target was a word that mismatched talker age. Results from the models fit to the data in the early region (Table 4.7) and the late region (Table 4.8) are presented below.

Table 4.7. Summary of Fixation Model #1 (data=early region, young and old target conditions)

Model: glmer (Looks to target ~ TalkerAge*WordAge + Time + TargetLocation +
(1+TalkerAge | Participant) + (1+TalkerAge:WordAge | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.483	0.371	-9.37	<.001
Talker age=young	0.946	0.568	1.67	.096
Word age=young	0.277	0.201	1.38	.169
Time	13.570	0.390	34.81	<.001
Target location=right	-2.088	0.096	-21.86	<.001
Talker age=young : Word age=young	**-0.519**	**0.469**	**-1.11**	**.268**

As shown in Table 4.7, listeners tended to fixate on the target more frequently during the early region when the talker was younger (*p*=.096) or when the target word was a young word (*p*=.169), but these effects did not reach significance. Time had a significant main effect and was included as a control variable; more fixations to the target was made as time proceeded (*p*<.001). A significant effect of target location indicates that targets were fixated less frequently when the target was located in the AOI on the right side than on the left side of the screen (*p*<.001). This effect is interpreted as a result of preference to look for the word from the left side (especially because the visual stimuli were written words). Such an effect of AOI location is not well established in the VWP literature, but Nixon et al. (2016) reports a similar effect even when using pictured objects. The interaction between talker age and word age indicates that listeners were less likely to fixate the target when talker age matched the word age, but this effect was not significant in the early region (*p*=.268).

Table 4.8. Summary of Fixation Model #2 (data=late region, young and old target conditions)

Model: glmer (Looks to target ~ TalkerAge*WordAge*Time + TargetLocation +
(1 | Participant) + (1+TalkerAge:WordAge | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.266	0.269	-12.13	<.001
Talker age=young	1.099	0.403	2.72	.006
Word age=young	0.565	0.174	3.25	.001
Time	12.307	0.261	47.16	<.001
Target location=right	-0.670	0.060	-11.14	<.001
Talker age=young : Word age=young	**-0.999**	**0.403**	**-2.48**	**.013**
Talker age=young : Time	-2.705	0.487	-5.56	<.001
Word age=young : Time	-1.124	0.242	-4.65	<.001
Talker age=young : Word age=young : Time	2.439	0.487	5.01	<.001

The model output fit to the data from the late region is presented in Table 4.8. Main effects of talker age and word age are significant, indicating that listeners fixated on the target more frequently when the talker was young (*p*<.01) or when the target was a young word (*p*<.01). Similarly with the early region (Table 4.7), time and target location had significant main effects in the late region (*p*<.001). Importantly, a significant interaction between talker age and word age was found; fixations to the target decreased when talker age and word age matched (*p*<.05).

This model included interactions with time. Time significantly interacted with talker age (*p*<.001) and word age (*p*<.001), indicating that the tendency to fixate the target more frequently when the talker was younger or when the target was a young word decreased as time proceeded.

A 3-way interaction was also significant; the tendency to look at the target less frequently when talker age matched word age decreased as time passed (*p*<.001).

The fixation analysis so far has shown that age incongruence led to more fixations to the target in general. But, the effect was not significant in the early region. It is inferable from Figure 4.5 that the mismatch-driven attention occurred steadily in both regions when the target was an old word, but young target trials induced different fixation behaviors between the early and late region; the early region of young target trials shows higher fixation rates for age-congruent talkers, while the pattern is reversed in the late region. I present the data from the young target condition (Section 4.4.2.1) and the old target condition (4.4.2.2) separately to see whether early fixations in young and old target trials were differentially influenced by age congruency.

4.4.2.1. The young target condition. Fixation data from the young target condition are plotted over time in Figure 4.6.

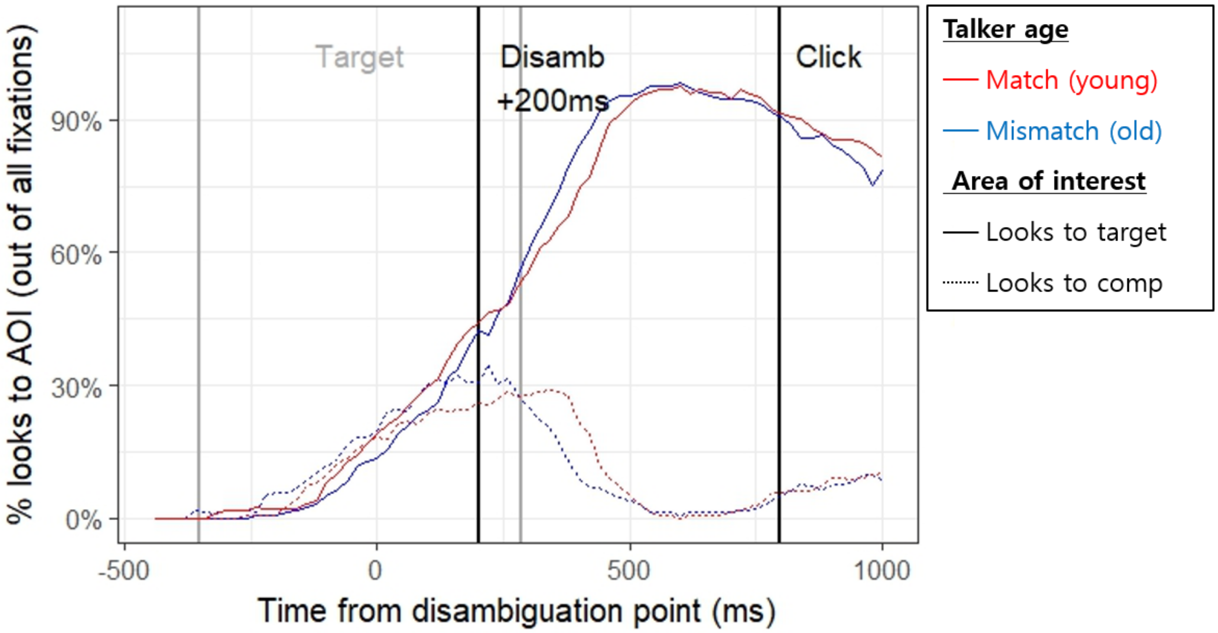


Figure 4.6. Proportions of looks over time (Condition=young target, competitor≠young)

The original prediction for the young target condition was that more target fixations would occur when produced by the age-congruent younger talkers than the older talkers. The predicted pattern is observed to a subtle degree in the early region in Figure 4.6; target fixation to young words is higher when talker age and target word age matched (red solid) than when they mismatched (blue solid). Also note that the mean target fixation rate diverged from the competitor fixation rate (indicated by the solid line diverging from the dotted line) at an earlier point for stimuli produced by younger talkers (red) than older talkers (blue). These patterns of fixation seem to indicate that the young target condition (unlike the old target condition) may have induced anticipatory looks to age-congruent targets before listeners perceived the disambiguating cue.

However, the pattern is reversed in the late region; target fixations occurred more frequently when produced by older talkers than younger talkers. The reversal of fixation frequencies upon the perception of the disambiguating cue indicates that auditory processing may have benefited from age-congruent phonetic detail while the signal was ambiguous; however, once the target was disambiguated, processing of the signal required greater attention for listeners to associate the acoustic signal with the pre-activated representation of an age-incongruent word (see, however, Section 4.5.2.1 for further discussion).

Statistical analyses were performed to examine whether the fixation patterns were significantly different between the early and late regions. First, models were fit to the data from the early region (Table 4.9) and the late region (Table 4.10), respectively.

Table 4.9. Summary of Fixation Model #3 (data=early region, young target condition)

Model: glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
(1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.812	0.426	-6.607	<.001
Talker age=young	**0.550**	**0.337**	**1.631**	**.103**
Time	12.514	0.477	26.234	<.001
Target location=right	-1.346	0.109	-12.349	<.001

Table 4.10. Summary of Fixation Model #4 (data=late region, young target condition)

Model: glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
(1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.570	0.320	-8.03	<.001
Talker age=young	**-0.344**	**0.176**	**-1.95**	**.051**
Time	10.467	0.318	32.92	<.001
Target location=right	-0.426	0.080	-5.30	<.001

Table 4.9 (early region) and Table 4.10 (late region) show main effects of time and target location; more fixations were made to the target (young words) as time proceeded (*p*<.001), and target fixations decreased when the target was located in an AOI on the right side (*p*<.001). The main effect of talker age in the late region is in the opposite direction from the effect in the early region. When a young target was produced by a younger talker (compared to an older talker), participants tended to fixate on the target more frequently in the early region (*p*=.103) and less

frequently in the late region (*p*=.051); however, neither effect reached significance[44]. The next model specifically tests whether the effect of talker age on fixations to young targets changed across the regions.

Another model was fit to the data from both regions to directly compare the fixation patterns in the two regions in a single model. This model included a binary distinction of region (early or late, centered), talker age, and their interaction as fixed effects. Time and target location were also included as control variables. To avoid collinearity between region and time, time in the late region was transformed to begin from 0.

Table 4.11. Summary of Fixation Model #6 (data=before click, young target condition)

Model: glmer (Looks to target ~ TalkerAge*Region + Time + TargetLocation +
(1+ Region | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-1.426	0.296	-4.83	<.001
Talker age=young	0.028	0.061	0.46	.647
Region=late	1.903	0.234	8.12	<.001
Time	10.778	0.256	42.09	<.001
Target location=right	-0.710	0.062	-11.44	<.001
Talker age=young : Region=late	**-0.585**	**0.123**	**-4.75**	**<.001**

---

[44] Looks to competitors were also tested (Fixation Model #5) on an assumption that competitor fixations would also be influenced by age-congruence as an additional effect, especially because competitors in the young target condition included old words (mean word age=0.77). However, no main effect of talker age was found in the early (β=-0.383, s.e.=0.547, *p*=.484) and late regions (β=0.360, s.e.=0.258, *p*=.163).

In Table 4.11, significant main effects are found for region (*p*<.001) and time (*p*<.001), indicating that participants looked at the target as time proceeded between the two regions and within each region. These two factors were not correlated (β=-0.098). Also, there is a significant interaction between talker age and region; when the word was produced by a younger talker (compared to an older talker), young word targets were fixated less frequently in the late region than in the early region (*p*<.001).

4.4.2.2. The old target condition. Fixation data from the old target condition are plotted over time in Figure 4.7.

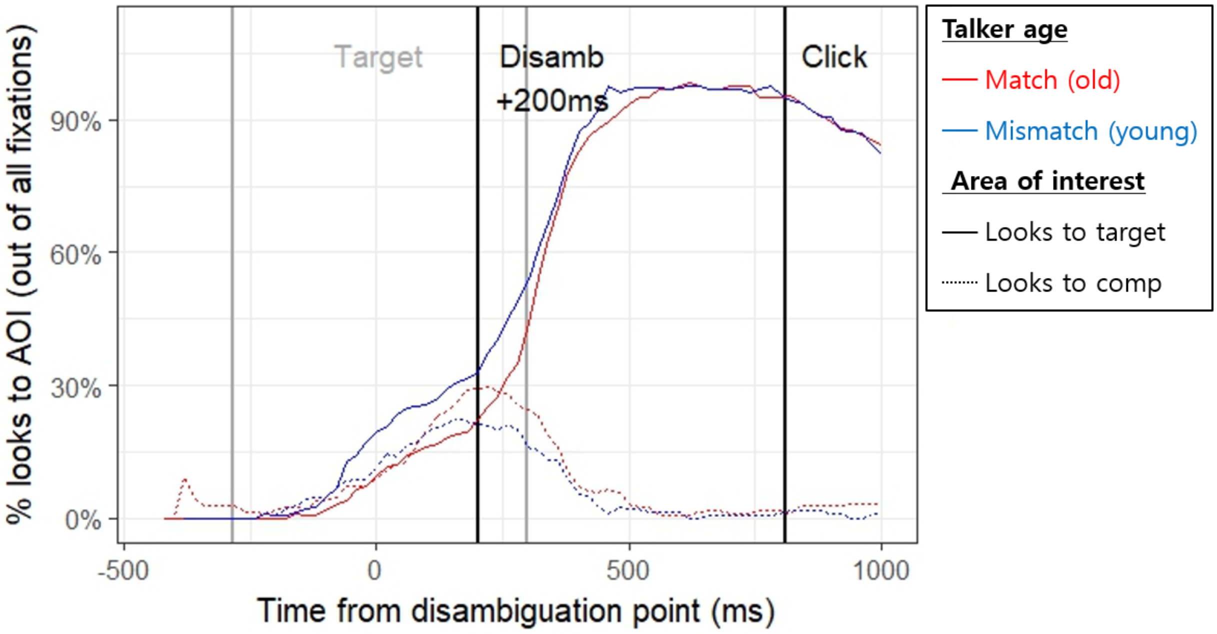


Figure 4.7. Proportions of looks over time (Condition=old target, competitor≠old)

Unlike in the young target condition, higher fixation probability for age mismatch is shown before the auditory signal was disambiguated. That is, old word targets appear to have been fixated more frequently when produced by age-incongruent talkers (younger) than congruent talkers (older) in both regions. The difference across talker age gradually decreased from the disambiguation+200ms point to the mean click time, indicating that mismatch-driven attention both arose earlier and waned earlier when an old word was heard than when a young word was heard.

Regression models were fit to the data from the old target condition to test the significance of higher fixation probability for old words produced by younger talkers. An interaction with time was included as a fixed effect in a model fit to the late region data, but not in the one fit to the early region. Table 4.12 and Table 4.13 present the outputs of the two models.

Table 4.12. Summary of Fixation Model #7 (data=early region, old target condition)

Model: glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
(1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-4.545	0.572	-7.941	<.001
Talker age=young	**1.706**	**0.604**	**2.825**	**.005**
Time	13.966	0.624	22.373	<.001
Target location=right	-2.782	0.159	-17.475	<.001

Table 4.13. Summary of Fixation Model #8 (data=late region, old target condition)

Model: glmer (Looks to target ~ TalkerAge*Time + TargetLocation +
(1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-4.292	0.311	-13.80	<.001
Talker age=young	**2.720**	**0.443**	**6.14**	**<.001**
Time	14.638	0.447	32.76	<.001
Target location=right	-1.059	0.091	-11.67	<.001
Talker age=young : Time	-6.406	0.867	-7.39	<.001

Results from both models show a significant main effect of talker age; target was fixated more frequently when the word was produced by a younger talker than when produced by an older talker. The effect size was greater in the late region (*p*<.001) than in the early region (*p*=.005), and so was the mean difference of fixation probability across talker age predicted by the model (indicated by the coefficients). Additionally, an interaction between talker age and time was significant in the late region; the tendency to look at the age-incongruent old target diminished gradually from the time listeners perceived the disambiguating phonetic cue to the time the mouse click was made (*p*<.001).

In the analysis of the age-related target so far, the fixation behavior in the early region appears to be different between young targets and old targets. Old words exhibited mismatch-driven attention while young words did not. To test the significance of the difference in fixation behavior, a model was fit to the target fixations in the early region of both conditions. Included as fixed effects in this model were main effects and an interaction of target word age (old or young) and talker-age congruency (whether talker age matched or mismatched word age), as well as time and target location.

Table 4.14. Summary of Fixation Model #9 (data=early region, young and old target conditions)

Model: glmer (Looks to target ~ WordAge* TalkerCongruency + Time + TargetLocation + (1+TalkerCongruency | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.964	0.323	-9.19	<.001
Word age=young	0.269	0.165	1.63	.103
Talker congruency=mismatch	0.134	0.191	0.70	.482
Time	11.997	0.347	34.59	<.001
Target location=right	-1.669	0.082	-20.49	<.001
Word age=young : Talker congruency=mismatch	**-0.647**	**0.075**	**-8.65**	**<.001**

The model output in Table 4.14 indicates a significant interaction between word age and talker-age congruency; when the target was a young word (compared to an old word) an age-incongruent talker led participants to fixate the target less frequently in the early region (*p*<.001). That is, the effect size of the incongruence-driven attention to the target in the early region was greater when there was an old word target.

This section (4.4.2) analyzed fixations made on target words that were manipulated with word age. As a summary of the results in the target-contrast conditions, age-incongruent voices induced more target fixations in the late region regardless of target word age, and the effect gradually waned as approaching the mouse click time. However, fixation behavior in the early region was different between the two conditions. While old word targets received more anticipatory looks when listening to an age-incongruent talker, young word targets did not exhibit this effect and trended in the opposite direction. In addition, the probability of fixation for young targets was significantly different between the two regions, indicating a reversal of the fixation behavior upon auditory disambiguation. Possible interpretations of these results will be provided in Section 4.5, with respect to listeners' prior experience and word age manipulations.

4.4.3. Fixation to age-related competitor

This section examines fixation patterns for young and old competitor words. The prediction for the competitor-contrast conditions was that listeners would look at an age-related competitor with a higher probability when an age-neutral target word was produced by a talker whose age matched the competitor's word age. In parallel, looks to targets would occur with a lower probability when there is age congruence between the competitor and the talker.

Fixation probabilities of these conditions are collapsed and plotted in Figure 4.8, with talker age and word age converted to match and mismatch. The match condition refers to the trials in which an age-neutral target word was produced by a speaker who matched the word age of the competitor.

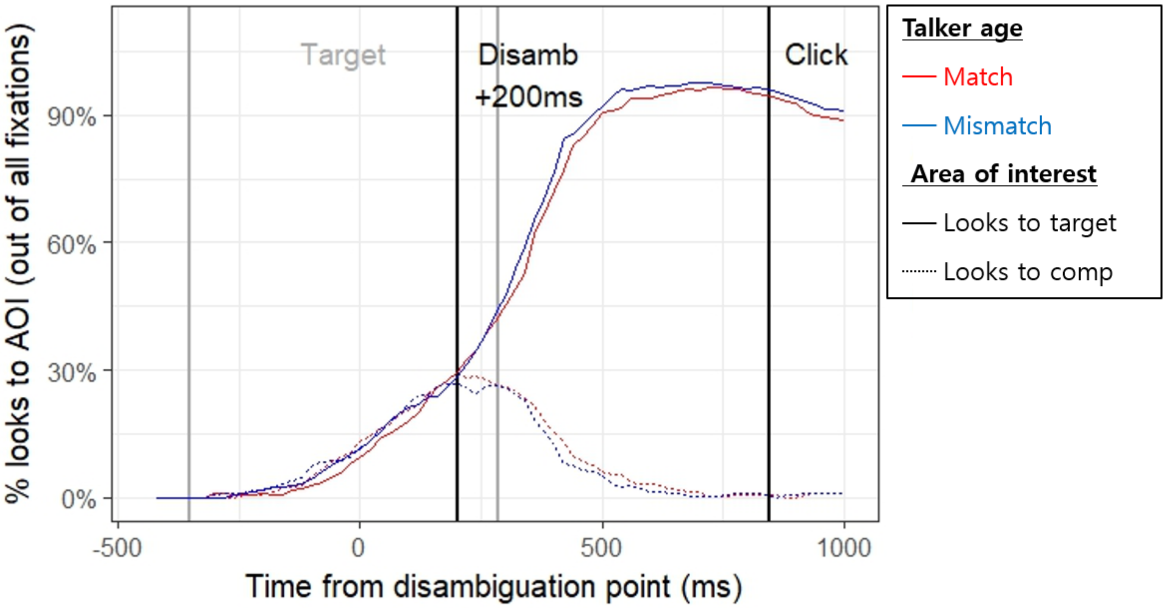


Figure 4.8. Proportions of looks over time (collapsing the young and old competitor conditions)

Fixations in the early region show little difference between conditions and AOIs, except that target fixation probability is slightly lower when the target word was produced by a talker with competitor-congruent age (red solid). In the late region, however, fixations to targets (solid) and competitors (dotted) exhibit slightly (but steadily) different probabilities between when talker age matched (red) and when talker age mismatched (blue). When listening to talkers with competitor-congruent age (compared to talkers with incongruent age), target fixation probability was lower while competitors show a higher probability. This indicates that age-congruent lexical candidates may have drawn listeners' attention even when they were not produced as a target. These patterns are also observed in Figure 4.9, in which competitor fixation probability is plotted by word age and talker age.

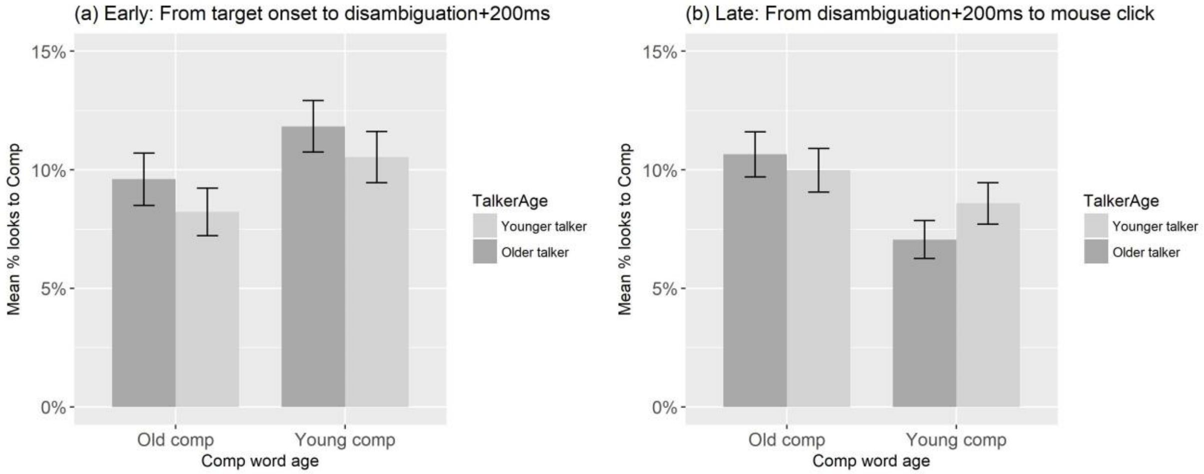


Figure 4.9. Mean competitor fixation probabilities by word age and talker age: (a) early region, (b) late region. Error bars represent 95% confidence interval of means by participants.

While there is little difference in competitor fixations across talker age in the early region (Figure 4.9 (a)), the late region (Figure 4.9 (b)) shows a slight tendency for participants to fixate the competitor more frequently when talker age matched the word age of the competitor (especially for young word competitors).

A model fit to the data from the early region of the competitor-contrast conditions (Fixation Model #10) showed no significant interaction between talker age and word age (β=0.080, s.e.=0.249, *p*=.748)[45].

As for the trend observed in the late region, note that the increase in fixations to age-congruent competitors in the late region is not incompatible with the results from the target-contrast conditions where an effect of incongruence-driven attention was found in the late region. In the competitor-contrast conditions, the age-related candidate was present in the scene but not produced as the target. Thus, there was no need for listeners to increase attention to the age-incongruent word (i.e., the competitor word) to match it with the disambiguated signal.

In contrast, if an age-related word is considered at an early point, we may expect that age-driven fixations will remain even after phonetic ambiguity is resolved. However, this effect was not supported by statistical analysis; in a model fit to the data from the late region (Fixation Model #11), an interaction between word age and talker age was not significant (β=0.412, s.e.=0.349, *p*=.238).

Target fixation in the competitor-contrast conditions was also analyzed, testing whether targets were fixated less frequently when the competitor was congruent with talker age. However, no significant interaction between talker age and word age was found in the early region (Fixation Model #12, β=-0.283, s.e.= 0.327, *p*=.388) and the late region (Fixation Model #13, β=-0.364, s.e.= 0.261, *p*=.164).

In order to examine whether different fixation patterns are observed between the young competitor condition and the old competitor condition, fixations are analyzed separately for the two conditions in Section 4.4.3.1 and 4.4.3.2.

---

[45] Model structure and outputs for models #10 through 25 are presented in Appendix III.

4.4.3.1 The young competitor condition. It was predicted in the young competitor condition that when the target was produced by a younger talker (compared to an older talker), more fixations to a young word competitor would be observed and fixations to an age-neutral target would be decreased. Fixation data from this condition are plotted in Figure 4.10.

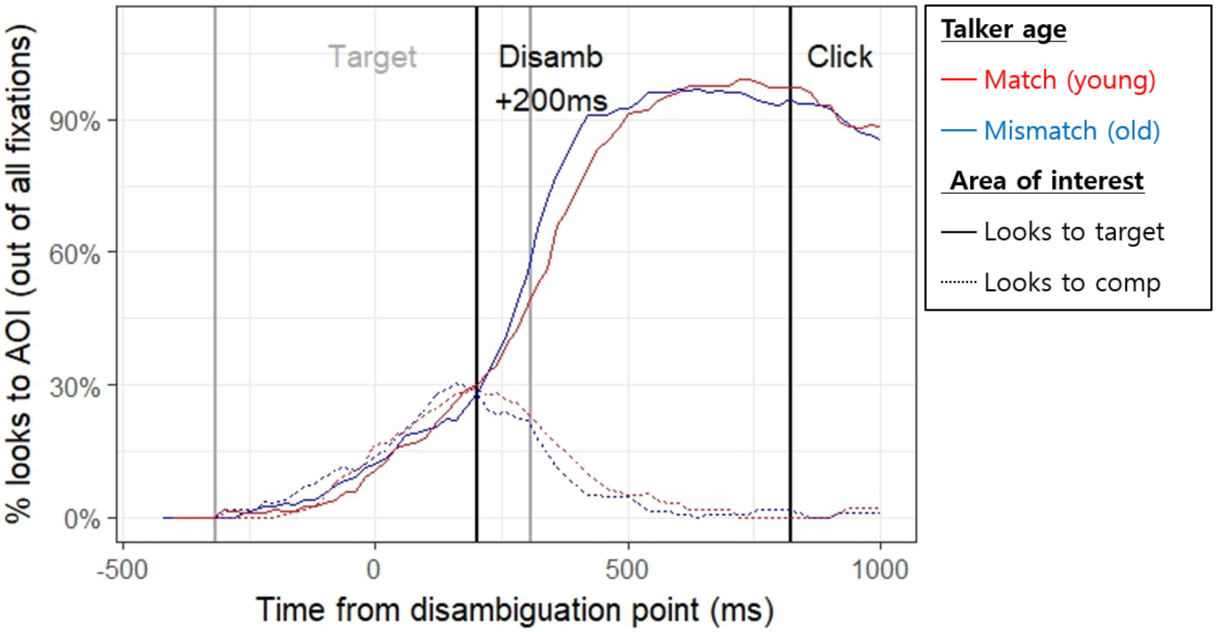


Figure 4.10. Proportions of looks over time (Condition=young competitor, target≠young)

While no steady patterns are observed in the early region, the late region shows the predicted patterns; young competitors were more frequently fixated (while neutral targets were less frequently fixated) when produced by younger talkers than older talkers. However, no evidence for these patterns was found in statistical analyses. A main effect of talker age did not reach significance, either in a model fit to fixations to competitor in the late region (Fixation Model #14, β=0.895, s.e.= 0.767, *p*=.243) or in a model fit to target fixations (Fixation Model #15, β=-0.496, s.e.= 0.360, *p*=.169). Fixations in the early region were also tested but no significant effects of talker age were found for looks to competitor (Fixation Model #16, β=0.059, s.e.= 0.363, *p*=.870) and target (Fixation Model #17, β=-0.032, s.e.= 0.650, *p*=.961).

4.4.3.2 The old competitor condition. The prediction for the old competitor condition was that when a neutral target was produced by an older talker (compared to a younger talker), more fixations to an old word competitor would be observed in the late region, accompanied by a decrease of target fixation. Data from this condition are plotted in Figure 4.11.

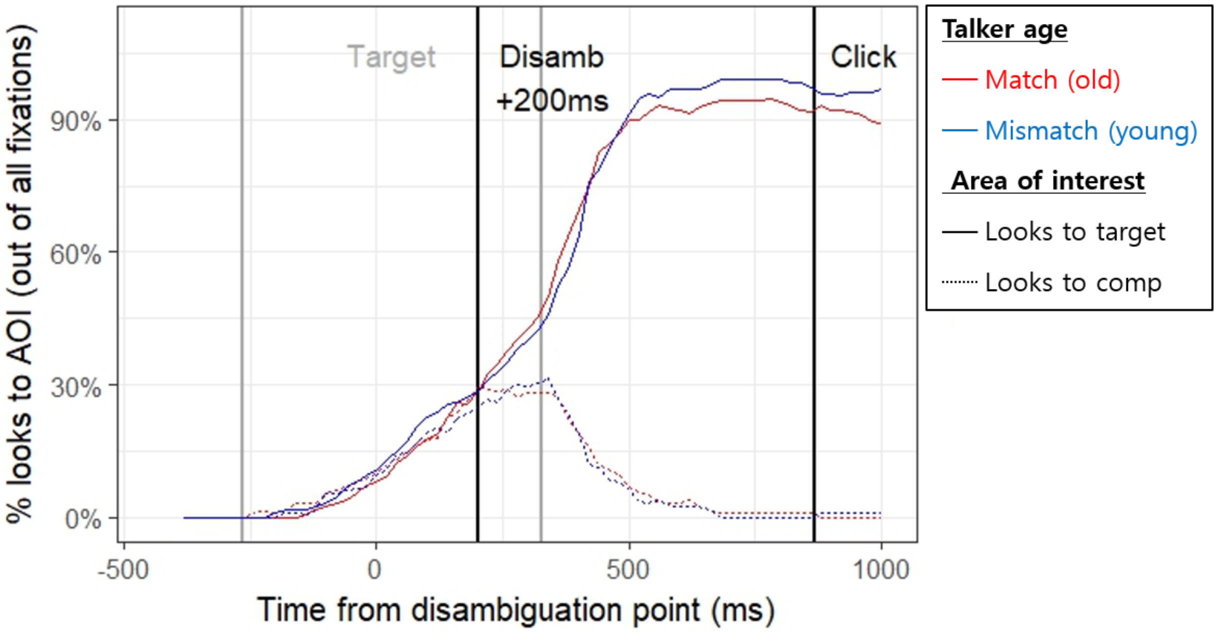


Figure 4.11. Proportions of looks over time (Condition=old competitor, target≠old)

Fixation probabilities in Figure 4.11 do not appear to support the predictions, and no effect of talker age was found in statistical analyses. In the early region, a main effect of talker age was not significant in models fit to fixations to competitor (Fixation Model #18, β=0.186, s.e.= 0.521, *p*=.720) and target (Fixation Model #19, β=0.272, s.e.= 0.714, *p*=.703). In the late region, talker age also did not influence fixations to competitor (Fixation Model #20, β=-0.095, s.e.= 0.255, *p*=.710) and target (Fixation Model #21, β=0.248, s.e.= 0.274, *p*=.366).

4.5. Discussion

To help understand the patterns found from each condition, a summary of the results is provided using bullet points.

- Mouse click RT
  - Target-contrast conditions: RT increased (i.e., slower target identification responses) when talker age is **congruent** (*p*=.083)
  - Competitor-contrast conditions: no effect of congruency between talker age and competitor age (*p*=.772)
  - All conditions collapsed: RT increased when talker age is **congruent** (*p*=.008)
- Eye fixations to young targets
  - Early region: more target fixations (i.e., greater attention driven to the target word) when talker age is **congruent** (*p*=.103)
  - Late region: more target fixations when talker age is **incongruent** (*p*=.051)
  - Reversal of fixation pattern upon disambiguation point (*p*=.001)
- Eye fixations to old targets
  - Early and late regions: more target fixations when talker age is **incongruent** in both early (*p*=.005) and late (*p*<.001) regions
  - No reversal of fixation patterns upon disambiguation point
- Eye fixations to young competitors
  - Early region: no difference
  - Late region: more competitor fixations (*p*=.243) and fewer target fixations (*p*=.169) when talker age is **congruent**

- Eye fixations to old competitors
  - No systematic trends in the either region

The results from the eye fixation analysis for the target-contrast conditions are generally consistent with the interpretation of the mouse click data, which I provided in Section 4.3.2, while fixation patterns in the competitor-contrast conditions are less clear. The overall results from Experiment 4 seem to point to an effect of preparatory attention driven to words produced by an age-incongruent talker. In the rest of this section, I will elaborate the observed preparatory effect with reference to the designs I used across experiments in this dissertation (Section 4.5.1) and possible explanations for the partial inconsistency with such interpretation (4.5.2). Then, I provide a summary of the implications along with pending problems and possible follow-up implementations (4.5.3).

4.5.1. Preparatory attention driven to social incongruence

Eye fixation analysis of the target-contrast conditions is compatible with the identification RT results, in which mouse click responses were facilitated when targets were produced by age-incongruent talkers (*p*=.008). When the two target-contrast conditions were collapsed, although the effect on eye fixations did not reach significance before the target word was disambiguated (*p*=.268), target words were fixated with a higher probability between the disambiguation point and the click response when the word was spoken by an age-incongruent talker (*p*=.013). Taken together with the result from identification RT, I interpret these results as an effect of preparatory attention during identification of a pre-activated lexical item.

The effect in the target-contrast conditions suggests that listeners strategically used available information in preparation for robust recognition of socially-incongruent speech signals. Since the lexical target was highly predictable, the task was oriented in the processing of lexical identification (rather than lexical access), in which the unfolding auditory cues were concurrently compared with the pre-activated representations. Given the talker information, listeners were also able to predict the stereotypical congruency between the words and the talker, or even predict how each candidate would sound when produced by the talker in the trial. In the context of experience-based models, expectation for the target word would increase, if the word had been previously encountered with phonetic realizations of speakers who are socially and/or phonetically comparable with the current talker.

However, since there were equal probabilities of the two candidates to be produced as targets, listeners began to pay greater attention to a socially-incongruent lexical item – for which listeners would have to compensate the lack experience with similar phonetic realization – as soon as the prior cues were perceived. When the attended word was heard, the attention was boosted in the later processing stage, during which the observed socially-incongruent signal was matched with the representations of the pre-activated lexemes. As a result, a word produced by an age-incongruent talker received more fixations (especially at the late region) and was eventually clicked on faster.

This effect was mainly found for old target words produced by (age-incongruent) younger talkers. Old targets were fixated more frequently both in the early region (p=.005), – indicating that the anticipation during phonetic ambiguity was influenced by attention readily drawn to the mismatching item – and in the late region (p<.001), indicating that the processing after disambiguation also required an additional processing cost to match the socially-incongruent signal with the pre-activated lexeme. The absence of the effect for young target words in the early region is not fully understood (although see 4.5.2.1), but the young target condition showed a reversed tendency in the late region where age-incongruent voice drew more fixations to the target (*p*=.051). Thus, the fixation patterns of young targets produced by older talkers is not incompatible with the facilitated click responses of these words.

The effect contrasts with the results reported in Chapters II and III, in which lexical recognition benefits from socially-congruent phonetic detail. In Experiments 1-3, uncertainty about the lexical item remained until the uniqueness point, at which all phonological cohorts were ruled out by auditory cues. Due to the uncertainty, the processing relied on the socio-indexical phonetic cues to facilitate access to the target lexeme. In contrast, during the process where pre-activated lexemes are identified, attention is readily paid to a socially-incongruent lexical item to recover the predicted lack of prior experience. The listeners' flexibility to counterplot incongruent social information highlights listeners' task-oriented adaptability to speech processing strategies.

4.5.2. Discrepancy in fixation behavior

The effect of mismatch-driven attention is not observed across the board. This section discusses two discrepant fixation patterns found across conditions and regions, which raise questions about the roles of preparatory attention. For example, to what degree would preparatory attention be strategically advantageous? Can young words and old words equally benefit from preparatory attention? When an age-incongruent candidate is attended, does the attention also increase the probability of the candidate to be the target? We discuss these questions with respect to the initially hypothesized age-congruence effect driven by socio-indexical phonetic cues.

4.5.2.1. Reversed patterns in the young target condition. The mismatch-driven attention is robustly found in the old target condition. However, the young target condition exhibited a transition of attention weight upon the point at which the phonetic ambiguity was resolved. Under phonetic ambiguity, young target words tended to be fixated more frequently when produced by age-congruent talkers (p=.103), while incongruent talkers drew more target fixations after disambiguation (p=.051). Although these trends were not significant in models separately fit to each time region, the talker age effect was significantly different across the two regions (p<.001), with a reversal of the fixation pattern. Further, in a model fit to the early region of the two conditions, the effect of age incongruence on fixation probability was significantly different between young word targets and old word targets (p<.001), indicating that the influence of age association on anticipatory fixations was not manifested in a uniform way between young and old target words. At least two possibilities can be considered to interpret the discrepancy.

First, since words produced by younger talkers (regardless of target word age) drew more target fixations than older talkers in the early region of the target-contrast conditions, there may be an advantage for younger voices that resulted in anticipatory attention to the target words. This might be possible since the target words were socially-salient lexical items (i.e., words strongly associated with age)[46] whereas the competitors were not as socially salient. However, we might then expect that early fixations would also have been affected by the competitor's strength of age-association. This is testable because the competitors in the target-contrast conditions included words with varying degrees of age-association. However, I found no such

---

[46] Because the signal was ambiguous during the early region, this advantage (if it existed) cannot be one that facilitates sub-lexical processing, like the main effect of talker age in previous chapters demonstrating that younger listeners are familiar with phonetic realizations of younger talkers.

evidence from further analysis[47]. Additionally, results from the competitor-contrast conditions are also inconsistent with this interpretation. Even though target items were age-neutral in these conditions, fixations to age-related competitors were not influenced by talker age in the early region (see Fixation Model #16 and #18).

Another possibility is that recognition of young word targets initially benefited from age-congruent phonetic cues before the phonetic disambiguation (as initially hypothesized) due to younger listeners' familiarity with the young words produced in the younger voice, but the influence of mismatch-driven attention overrode the age-congruence effect in the later processing stage where listeners eventually interpreted the socially-incongruent realization as the target. In contrast, old word targets may not be able to enjoy the advantage of phonetic congruence due to insufficient experience. Because younger listeners are less familiar with speech produced by an older talker and with old-associated lexical items (as evidenced in Experiment 1 and 2), representations of old words in the participants' lexicon may not be as strongly encoded with age-congruent phonetic detail as representations of young words are.

---

[47] Multiple models were fit to fixations in the early region of the target-contrast conditions to test for such an effect. For example, in models where fixations were predicted by main effects and interaction of talker age and competitor strength (a binary distinction of whether an absolute value of word age is smaller or greater than 0.7 (weak or strong association)), the interaction did not influence either target fixation (Fixation Model #22, $\beta=0.021$, s.e.=0.146, $p=.884$) or competitor fixation (Fixation Model #23, $\beta=0.121$, s.e.= 0.132, $p=.362$).

Thus, old-associated phonetic cues may not have been weighted sufficiently to draw attention to an (age-congruent) old word candidate beyond the attention driven to a (age-incongruent) young word candidate[48]. This interpretation may be more consistent with predictions of experience-based models than the first interpretation but is not supported by the data, either. If young words drew more early fixations due to age congruence of the phonetic information, this effect may also be observed for young competitor words used in the old target conditions; however, I found no evidence from further analysis[49], leaving the cause of the discrepancy as an open question.

However, it is also notable that the congruence-induced anticipatory looks to young target words in the early region was not observed in a previous version of this experiment[50], in which the second slide (fixation cross) was not present. In that experiment, without intervention of focal point fixations, attention driven to mismatch during the frame sentence was manifested as a drastic increase in the target fixation rate in the early region for both young and old target words produced by age-incongruent talkers, lasting until the mouse click. Thus, it appears in the current experiment that preparatory attention that had been drawn to an incongruent item while listening to the frame sentence was reset as soon as the incongruent signal was heard since

---

[48] This interpretation does not necessarily imply that recognition of old words would not benefit from older voice (which contradicts with the findings in the previous chapters). Even if old words realized in older voice are not densely clustered in memory, socially-salient representations may benefit from associations at a higher level as discussed in Chapter I.

[49] In models fit to the old target condition, an interaction between talker age and competitor word age (binary) did not significantly influence fixation probability of the target (Fixation Model #24, β=-0.061, s.e.=0.123, *p*=.622) and competitor (Fixation Model #25, β=-0.154, s.e.=0.122, *p*=.208) in the early region.

[50] I am not reporting the methods and results from that experiment in more detail because its design differed in multiple ways making it difficult to directly compare its results with the current data.

listeners focused on the fixation cross. This suggests that the discrepant pattern found in the early region of the young target condition should be considered as a rapid influence of the target signal, although it may be disputable whether the effect was indeed driven by phonetic congruence between the realization and the representation, and why the effect was not observed for old target words.

4.5.2.2. Fixations in the competitor-contrast conditions. The second discrepancy is found in the competitor-contrast conditions, in which there is no indication of mismatch-driven attention in both regions[51]. If the age-related competitor had drawn preparatory attention when it mismatched talker age, it may be possible that age-incongruent competitors (especially old competitors produced by younger talkers) are more frequently fixated in the early region. But, the early region showed no effect of age associations. From a similar viewpoint with the second interpretation in Section 4.5.2.1, the null effect on early fixations seems to be due to the presence of age-neutral target words with high lexical frequency. Since listeners are expected to have sufficient prior experience with these frequent words, they can form an expectancy about the phonetic realizations regardless of talker age. Thus, even if age-incongruent competitors had drawn preparatory attention, it is possible that the attention was neutralized by an experience-based effect for phonetically-rich representations of the neutral targets.

On the other hand, if there had been an effect of mismatch-driven attention in the late region, the effect would have been manifested in a way that fixations already drawn to an age-incongruent competitor in the early region remained, delaying subsequent target fixations.

---

[51] As noted earlier, unlike in the target-contrast conditions, mismatch-driven attention in the competitor-contrast conditions would be observable (if it existed) in fixations to competitors (not targets), because targets in these conditions were lexical items that are frequently used by all age groups.

However, an opposite trend was observed; competitors tended to be fixated with a higher probability when the target was produced by an age-congruent talker (as initially hypothesized)[52], although the effect did not reach significance in all models fit to these conditions. This suggests that the attention did not boost the probability of encountering the incongruent candidate to a degree that it prevented interpretation of the auditory stimulus. However, when the signal informed that the incongruent candidate was the target, mismatch-driven attention induced higher attention and facilitated the response.

4.5.3. Summary of implications and pending problems

The question I raised in the beginning of the chapter – whether anticipatory fixations during phonetic ambiguity is influenced by phonetically-rich memories – remains unsolved. Such an effect, if any, would have provided evidence for rapid and concurrent integration of socially-indexed phonetic realization. Or, more generally in a psycholinguistic context, such an effect would have implications for the temporal aspects of influx of lexical information.

The absence of such evidence may be attributable to some methodological decisions I made. First, the presence of top-down social information (i.e., pre-activation of prior cues) and the use of socially-salient lexical items (rather than words associated with age in distribution but not in stereotypes) seem to have provided excessive information that highlighted the possibility for age-incongruent combinations. Second, the use of competitor words that are associated with the opposite age groups may have enhanced the mismatch-driven attention. Third, fixation rates were dominantly influenced by the AOI location, especially because only two orthographic

---

[52] It is also notable that the trend was mainly found for young competitors, with which listeners have sufficient experience.

lexical competitors were used. In fact, there are compounding factors that may need to be further examined in the context of the current experiment. The strong influence of AOI locations, for example, may have overridden some effects of the predictors I tested, so it would be worth examining fixation patterns while holding AOI location constant.

Changes to the experimental design and items are needed to reexamine whether the influence of probabilistically congruent information between the talker and the word can arise as soon as the auditory input begins to unfold, while contextual cues other than the phonetically-encoded information are controlled for. Research addressing these issues would help expand our knowledge about the online processing of socially-conditioned lexemes.

On the other hand, interpretation of the results in the context of preparatory attention reveals listeners' ability to cope with socially incongruent combinations of talker and lexical information, especially when processing unfamiliar lexical items (i.e., old words), which lack phonetic encoding. This interpretation is consistent with experience-based models in that listeners make full use of contextual information, (1) to prepare for socially incongruent combinations (manifested in the early region of the old target condition), and (2) to increase attention for matching the signal to a socially-incongruent lexical item (manifested in the late region of the young target and old target conditions). Listeners used not only the high predictability of the lexical item but also higher-level information, including abstract age-related associations and probabilistic fitness of the given talker's voice with candidate representations. In this vein, it would also be worthwhile to examine how pre-activated social information is mediated by unfolding phonetic cues during recognition of socially-conditioned speech.

CHAPTER V

General Discussion and Conclusion

In this chapter, I discuss two main implications of this dissertation in light of experience-based models. In Section 5.2, I argue that socially-indexed phonetic forms and words are closely tied in the lexicon and cognitive processing, so that socio-indexical phonetic cues not only lead listeners to adjust expectations for subsequent speech but also can index associated lexemes rapidly and directly. In Section 5.3, I underscore the role of stereotypical associations in processing of socially-salient lexical items and discuss how the additional effect of higher-level associations can be accounted for in existing experience-based models. Prior to the discussion of the implications, I present a brief summary of the results of the four experiments presented in this dissertation in Section 5.1.

5.1. Summary of results

The two lexical decision experiments presented in Chapter II replicated the age-congruence effect (Walker & Hay, 2011) using Korean, demonstrating that lexical recognition was improved when the word and the talker were indexed to the same age group. Consistent with Walker and Hay's results, the effect was observed when the associations were determined by the unbalanced usage frequencies of words over social categories; however, the effect was enhanced when words were stereotypically linked to age (Experiment 1). Further, the effect occurred even when listeners held no expectations about the talker before the word onset (Experiment 2); the effect size was not different across the two experiments.

The phonetic priming experiment in Chapter III (Experiment 3) demonstrated that exposure to a single phonetic variable can influence predictability for a subsequent lexical item and thus facilitate recognition of words indexed to the age that the phonetic variant is associated with. Also investigated was whether activation of social information was necessary to see an effect. While the effect was significant for the subgroup of participants who perceived age difference of the talkers, the effect did not reach significance for those who did not, indicating that activation of social information may have played a role. However, it is also possible that listeners benefited from the direct links between sociophonetic variants and age-indexed words. First, the effect was only observed for lexical items immediately preceded by an age-related phonetic variant, but not after a variant immune to age even though it was produced by the same talker. Second, the difference in effect size between listeners who perceived the difference in age and those who did not was not significant. Last, although reaction times were affected by the primes, accuracy rate was not – unlike in Experiments 1 and 2 – indicating that the observed effect may have been drawn by involuntary attention to the subtle phonetic cue.

In the lexical identification experiment in Chapter IV (Experiment 4), lexical targets that were socially incongruent with the talker information drew more frequent eye fixations and resulted in faster identification than congruent words. Since the age-related associations between words and the talker were predictable with the prior cues, I interpreted incongruent condition advantage as evidence that attention to the age-incongruent target word was increased during the time region in which the pre-activated lexical representation was matched with the disambiguated signal. For old word targets, the effect was also robust in the early region before disambiguation, suggesting that a greater degree of preparatory attention was drawn to words with lower familiarity to compensate the small number of phonetic memories for these words.

5.2. Socially-conditioned interplay between phonetic realizations and words

Building on Walker and Hay's (2011) finding that lexical access is facilitated when age-indexed information of the talker and the word is congruent, the results from this dissertation provide additional information about how socially-indexed words and phonetic forms are encoded in the speech processing system. This section discusses the implications of the results in relation to the way the interplay between phonetic realizations and words are modelled in the experience-based framework.

As discussed in Chapters II and III, there are two possibilities regarding the status of social information in triggering the age-congruence effect. First, socio-indexical information in the phonetic cues affects lexical processing because it changes listeners' beliefs about the talker. In other words, the voice could activate social information as a connecting link between words and sounds, leading to activation of lexical items that are indexed to social information also indexed with the activated phonetic information. Alternatively, the relation between phonetic forms and lexemes may be more closely encoded in cognition and in processing routines (e.g., socially-informative phonetic detail is encoded in socially-indexed lexical representations), so that phonetic realizations influence lexical processing rapidly (i.e., immediately upon retrieving the signal) and directly (i.e., not via abstract social information). Both aspects are demonstrated across experiments reported in the previous chapters.

Under the first interpretation, the age-congruence effect is understood, by exemplar models, as a top-down process drawn by age-related exemplars that are readily activated prior to the exposure to the target. Likewise, in the Bayesian framework, inferences are influenced by expectations about the talker with reference to prior probability (i.e., implicit knowledge about the covariation of phonetic cues and word usage). In line with these predictions, this dissertation shows that the contextually-encoded social information manipulated listeners' expectations prior to the target signal, and listeners efficiently utilized the expectancy according to the purpose of the task. For example, building on previous work showing that implicit exposure to non-linguistic prime for top-down social information can trigger socially-conditioned phonetic processing (Hay et al., 2006a; Hay & Drager, 2010), Experiment 3 demonstrated a signal-based effect, in which listeners formed expectations for lexical items based on exposure to a single phonetic variant and the expectations induced difference in processing speed. While this effect highlights listeners' ability to exploit socio-indexical cues during lexical access under uncertainty, Experiment 4 demonstrated the ability to selectively adjust the context-driven expectation in line with the purpose of the task to overcome processing difficulty imposed by pre-activated social incongruence.

As for the second interpretation of the age-congruence effect, the close relationship between phonetic realizations and socially-indexed lexemes is well defined in exemplar models, in which memories of socially-informative phonetic detail are encoded in the lexical representation, and recognition is facilitated when phonetic detail in the signal and the representation match. Walker and Hay (2011) interpret their results in line with this prediction, and some results from this dissertation can be interpreted to support this view, outlined below.

First, it is inferable from the comparable effect size of Experiments 1 and 2 that, within the context of a lexical decision task, influences of socio-indexical phonetic cues are not boosted by prior predictability for talker identity but instead occur immediately as phonetic cues are retrieved. This result is consistent with exemplar-based models in which lexical representations that contain phonetic detail that matches the signal are activated most quickly and strongly. However, it is also possible that talker-appropriate exemplars were activated by the signal rapidly (i.e., during the auditory presentation of the word), and then activation spread to associated lexemes.

More evidence that the talker representation does not need to be activated during the process comes from Experiment 3. Results from Experiment 3 suggest that exposure to a socially-indexed phonetic cue boosted activation rate of socially-congruent lexemes either via activation of social information or via direct indices between them, or both. If future experiments replicate this effect while teasing apart confounding factors (e.g., whether listeners were aware that there were two female talkers), it would provide stronger evidence that explicit expectations about, or even awareness of, the talker characteristics play little role beyond the effect drawn by perceptual congruence of the signal with prior experience with phonetic realizations of the word. I believe future investigation using an online method will also provide further insight into this question (e.g., testing whether the very initial stage of lexical processing is affected in real time, or testing whether talker characteristics is necessarily attended to), as well as the question about the temporal aspects of top-down integration during socially-conditioned speech processing and spoken word recognition in general (see Section 1.1). Consequently, pursuing the line of research introduced and reported in this dissertation will help verify the claim that lexical access is concurrently consulted by acoustic congruence between the signal and the representation, as well as contextual congruence.

Additionally, the effect of acoustic match is also compatible with Bayesian models, which do not assume socially-encoded representations but instead posit rapid and flexible adaptation of the generative model. In such a mechanism, listeners continually update their beliefs about the talker-appropriate model based on prior experience with similar-sounding talkers' models (Kleinschmidt & Jaeger, 2015). However, as discussed in Chapter III, the direct interplay between phonetic and lexical information is more explicitly specified in connectionist models (e.g., exemplar models). Since Kleinschmidt & Jaeger’s Bayesian adaptation to the talker-appropriate generative model relies on the talker representation, its prediction may not be consistent with my interpretation of Experiment 3 that the age-congruence effect occurred even though the priming was not salient, or strategically reliable, enough to influence the perceptual representation of the talker. However, since social characteristics of words and phonetic variants can be encoded independently as factors contributing to prior probability of the word, either of them will be able to inform the other based on their covariance in the Bayesian framework.

5.3. Socially-salient indexical information

The experiments reported in this dissertation found robust effects of words stereotypically associated with age. In Experiment 1, I specifically demonstrated that the effect of age-indexed phonetic realizations on lexical recognition is enhanced when words are stereotypically linked to age and suggested that indices formed based on sociophonetic distributions can be reinforced by stereotypes. The additional role of stereotypes may appear inconsistent with Walker and Hay (2011), in which the effect of relative word frequency was present whereas stereotypes did not show an effect. This section presents further information about the two different aspects of social information indexed to lexical items with reference to the methods used across the studies, and discusses how existing models of speech perception can be modified to account for both aspects.

The different results can be interpreted to have arisen from differences in methods targeting different aspects of perceptual associations. Since Walker and Hay selected the stimuli based on relative frequency data, the age distinction in their lexical stimuli was less likely to emerge above a conscious level of association. Thus, what they found is that words' socio-indexical distribution affects lexical recognition automatically without listeners' explicit awareness of the association, which in itself is a different type of evidence for the acoustic-match effect. In contrast, lexical items used in my experiments were selected based on word-stereotypes of native speakers. Exposed to words on a wide spectrum of age-related variability, listeners were more likely to make use of the age-contrastive information indexed to the lexical stimuli. As a result, this dissertation finds that stereotype is an additional predictor of socially-conditioned lexical access, the effect of which appeared to be stronger than an effect of distribution when appropriate items are used.

The robust effect of word-stereotypes calls for further research on how socially-salient indexical information is encoded in speech processing. As discussed in Chapter I, indices between linguistic forms and social information do not seem to be weighted entirely by usage frequencies over social categories. When the phonetic variant is socially salient, its perception is influenced by stereotypes about the way people realize the variable, even when the form is indeed rarely found in speech of the community (Niedzielski, 1999; Sumner et al., 2014). However, while existing models in the experience-based framework predict perception behavior primarily based on statistical calibration of phonetic-level usage frequency, the role of listeners' conscious knowledge about social and ideological indexicality in a given social context (e.g., word-stereotypes) remains largely underspecified.

Since stereotypical links are formed at a higher level of awareness (Labov, 1972), they may need to be specified independently in cognitive models of language processing, although further empirical evidence is needed. As discussed in Section 2.3, the hybrid models (Pierrehumbert, 2002, 2003, 2016; Pitt, 2009; Pinnow & Connine, 2014) allow generalizations of experienced phonetic patterns into abstract lexical representations, but these processes also rely on exposure frequency for the generalization process, not specifying a separate route for indices created and updated by social salience. Exemplar-based models need to account for effects arising from socially-salient tokens by, for example, using multi-layered structure of associations or including boosted activation triggered by encoding strength. Likewise, Bayesian models can be readily adjusted to account for the expectation for lexical items by parameterizing lexical salience or talker-word relations within the social context as prior probability. In contrast, Sumner et al.'s (2014) dual-route mechanism addresses the problem of disproportional emphasis on token frequency, by positing that phonetic and social information continuously inform the other during lexical access, and socially salient tokens (e.g., canonical forms) are socially weighted even if they are infrequent. Results from this dissertation suggest that their models can be developed to place social weighting for a wider range of social meanings (including stereotypes).

5.4. Concluding remarks

Previous research has established the legitimacy of the experience-based view in pursuing questions about recognition of socially-conditioned speech. This is in line with the growing views of psycholinguists who emphasize the roles of subcategorical acoustic detail. Exemplar- and Bayesian-based models differ in some basic assumptions in lexical processing – e.g., how lexemes are represented (variable or abstract) or how they are accessed (activation-based or rule-based) – and it remains a challenge for psycholinguists and sociophoneticians to determine to what degree indexical information is abstracted in linguistic representations, or more radically, whether the structural relationship between social information and linguistic units is encoded at a representational level or parameterized by flexibility of access routes. However, both approaches put forward a common understanding that linguistic cognition is shaped by an individual's lifetime-experience with socially-conditioned linguistic properties.

In this respect, the overall results of this dissertation are consistent with their predictions for word processing under the influence of sociophonetic and lexical variation. Through a series of perception experiments, I demonstrated that listeners can strategically use socially-conditioned links between phonetic realizations and words. The results suggest that phonetic- and lexical-level memory and processing are jointly shaped by statistically meaningful distributions of experienced phonetic forms and are refined by socially meaningful experiences with salient tokens. In addition, this dissertation contributes to expanding the scope of sociophonetic research – which used to be mainly concerned with the interplay between phonetic forms and indexical information – highlighting the trigonal relationship where indexical properties of lexical items also play a role. I believe continuing research in this vein will bring a deeper appreciation of the experience-based cognitive system intertwined with a wider range of linguistic and social factors.

APPENDIX I

Lexical stimuli in Experiments 1-3

Lexical stimuli (real word items only) used in Experiment 1-3 are listed in an ascending order for word age (ST score). Filler items are not assigned word age. For items for which non-standard phonetic form is associated with age, Hangul transcription is based on surface forms, rather than standard orthography. Phonemic transcription into Roman characters is based on “The Yale Romanization System”. All items listed below (N=384) are used in Experiment 1 and 2. Items used in Experiment 3 are indicated by “used” in the column “Exp3”. The stimuli are available for download from: https://www.bloomsbury.com/cw/experimental-research-methods-in-sociolinguistics/.

Item #	Word	Phonemic transcription	ST score	UA score	Exp3
1	꿀잼	kkwulcaym	-1.95	-2.03	Used
2	노답	notap	-1.95	-2.27	Used
3	개드립	kaytulip	-1.95	-2.48	Used
4	엄빠	emppa	-1.94	-1.80	Used
5	인적성	incekseng	-1.94	-1.41	Used
6	짤방	ccalpang	-1.89	-1.79	Used
7	노잼	nocaym	-1.89	-2.05	Used
8	득템	tuktheym	-1.89	-0.87	Used
9	김치녀	kimchinye	-1.89	-1.05	
10	에이쁠	eyippul	-1.89	-1.38	Used
11	꿀벅지	kkwulpekci	-1.89	-0.69	
12	취준생	chwicwunsayng	-1.89	-1.40	Used
13	페친	pheychin	-1.89	-1.26	Used
14	리트윗	lithuwis	-1.89	0.04	Used
15	된장녀	toyncangnye	-1.89	-0.86	

16	낮져밤이	naccyepami	-1.89	-1.01	Used
17	레알	leyal	-1.84	-1.47	Used
18	멘붕	meynpwung	-1.84	-0.89	Used
19	지못미	cimosmi	-1.84	-0.44	Used
20	페이스북	pheyisupwuk	-1.84	-0.46	Used
21	개강총회	kaykangchonghoy	-1.84	-1.25	Used
22	꽈잠바	kkwacampa	-1.84	-1.80	Used
23	떡실신	tteksilsin	-1.84	-1.26	
24	병맛	pyengmas	-1.83	-1.34	Used
25	팀플	thimphul	-1.79	-2.32	Used
26	룸메	lwummey	-1.79	-1.57	Used
27	스펙	supheyk	-1.79	-0.78	Used
28	찐따	ccintta	-1.79	-1.33	Used
29	깜놀	kkamnol	-1.79	-1.58	Used
30	열폭	yelphok	-1.79	-1.44	Used
31	밀당	miltang	-1.79	-1.05	Used
32	광탈	kwangthal	-1.79	-1.65	Used
33	고딩	koting	-1.79	-1.18	Used
34	넘사벽	nemsapyek	-1.79	-1.31	Used
35	개념녀	kaynyemnye	-1.79	-0.31	
36	자소서	casose	-1.79	-1.70	Used
37	유튜브	yuthyupu	-1.79	-0.41	Used
38	파스타	phasutha	-1.79	-0.63	Used
39	토익	thoik	-1.74	-1.01	Used
40	빡침	ppakchim	-1.74	-2.43	Used
41	베프	peyphu	-1.74	-1.52	Used
42	해외직구	hayoycikkwu	-1.74	-0.58	Used
43	모태솔로	mothaysollo	-1.74	-1.53	Used
44	룸메이트	lwummeyithu	-1.74	0.01	Used
45	트위터	thuwithe	-1.74	-0.21	Used
46	솔까말	solkkamal	-1.74	-0.80	Used
47	기레기	kileyki	-1.71	-0.66	
48	겨털	kyethel	-1.68	-1.81	Used
49	라떼	lattey	-1.68	-0.56	Used

50	안습	ansup	-1.68	-0.89	Used
51	훈남	hwunnam	-1.68	-1.04	Used
52	썩소	ssekso	-1.68	-1.29	Used
53	뽀대	ppotay	-1.68	-0.33	Used
54	먹튀	mekthwi	-1.68	-0.63	Used
55	듣보잡	tutpocap	-1.68	-1.62	Used
56	귀요미	kwiyomi	-1.68	-1.14	Used
57	피드백	phitupayk	-1.68	-0.55	
58	학고	hakko	-1.67	-1.03	Used
59	파티셰	phathisyey	-1.67	-0.11	
60	눈팅	nwunthing	-1.63	-1.19	Used
61	강추	kangchwu	-1.63	-0.42	Used
62	동기화	tongkihwa	-1.63	-1.16	Used
63	와이파이	waiphai	-1.63	-0.38	Used
64	폭망	phokmang	-1.60	-1.11	
65	놀토	noltho	-1.58	-0.25	Used
66	랩탑	laypthap	-1.58	-0.45	Used
67	돌직구	tolcikkwu	-1.58	-1.00	
68	귀차니즘	kwichanicum	-1.58	-0.90	
69	품절남	phwumcelnam	-1.58	-0.75	
70	소개팅	sokaything	-1.58	-0.96	
71	친추	chinchwu	-1.56	-1.86	
72	킹왕짱	khingwangccang	-1.56	-0.46	
73	볼매	polmay	-1.53	-1.49	
74	열공	yelkong	-1.53	-1.45	Used
75	얼짱	elccang	-1.53	-0.40	Used
76	대박	taypak	-1.53	-0.93	
77	초딩	choting	-1.53	-0.51	
78	어플	ephul	-1.53	-1.07	Used
79	엄친아	emchina	-1.53	-0.62	Used
80	프로젝트	phuloceykthu	-1.53	-0.43	
81	깔맞춤	kkalmacchwum	-1.53	-0.72	
82	디카	tikha	-1.50	0.27	
83	치맥	chimayk	-1.47	-1.35	

84	동아리	tongali	-1.47	-0.73	
85	카푸치노	khaphwuchino	-1.47	0.35	
86	본방사수	ponpangsaswu	-1.47	-0.47	
87	바리스타	palisutha	-1.47	-0.70	
88	몸짱	momccang	-1.42	-0.09	Used
89	쌩얼	ssayngel	-1.42	-1.28	
90	킹카	khingkha	-1.42	-0.13	
91	노트북	nothupwuk	-1.32	-0.23	Used
92	쇼핑몰	syophingmol	-1.32	-0.28	
93	요거트	yokethu	-1.21	-0.62	Used
94	공돌이	kongtoli	-1.11	-0.96	
95	꼰대	kkontay	-1.05	-0.57	
96	환타지	hwanthaci	-1.00	0.39	Used
97	프로덕션	phuloteksyen	-0.95	-0.08	Used
98	스파게티	suphakeythi	-0.95	-0.17	Used
99	프린터	phulinthe	-0.95	-0.20	Used
100	안구인식	ankwuinsik	-0.93	0.02	Used
101	일인자	ilinca	-0.89	0.13	Used
102	면접비	myenceppi	-0.76	-0.19	Used
103	뺑뺑이	ppayngppayngi	-0.76	-0.02	Used
104	지문인식	cimwuninsik	-0.63	-0.27	Used
105	공순이	kongswuni	-0.50	0.43	Used
106	화일	hwail	-0.38	0.64	Used
107	추리닝	chwulining	-0.26	-0.56	Used
108	캬라멜	khyalameyl	-0.22	0.27	Used
109	크레파스	khuleyphasu	-0.21	0.44	Used
110	자유분방	cayupwunpang	-0.21	0.69	Used
111	히로뽕	hiloppong	-0.16	0.18	Used
112	유대감	yutaykam	-0.11	0.29	Used
113	그저께	kucekkey	-0.11	-0.25	Used
114	몸짓	momcis	-0.11	0.28	Used
115	빵꾸	ppangkkwu	-0.05	-0.41	Used
116	알레르기	alleyluki	-0.05	0.01	Used
117	꼭대기	kkoktayki	-0.05	-0.14	Used

118	손짓	soncis	0.00	0.83	Used
119	여보	yepo	0.00	0.39	Used
120	후까시	hwukkasi	0.00	0.09	Used
121	강냉이	kangnayngi	0.05	0.18	Used
122	와이프	waiphu	0.05	1.17	Used
123	손가락질	sonkalakcil	0.05	-0.16	Used
124	쪼꼬렛	ccokkoleys	0.05	0.04	Used
125	빽미러	ppaykmile	0.06	0.56	Used
126	구제불능	kwuceypwulnung	0.11	0.27	Used
127	이념	inyem	0.11	0.91	Used
128	바란스	palansu	0.14	0.88	Used
129	분풀이	pwunphwuli	0.16	-0.12	Used
130	날벼락	nalpyelak	0.21	0.51	Used
131	까닭	kkatalk	0.21	0.27	Used
132	잠바	campa	0.21	0.26	Used
133	늘상	nulsang	0.21	0.14	Used
134	품위	phwumwi	0.21	0.71	Used
135	바가지	pakaci	0.22	0.28	Used
136	어저께	ecekkey	0.22	0.14	Used
137	바느질	panucil	0.26	0.42	Used
138	조끼	cokki	0.26	0.42	Used
139	빳데리	ppasteyli	0.26	-0.45	Used
140	바통	pathong	0.27	0.58	Used
141	빠꾸	ppakkwu	0.29	0.01	Used
142	식구	sikkwu	0.32	0.64	
143	고향	kohyang	0.32	0.41	Used
144	후라이판	hwulaiphan	0.32	0.57	Used
145	형님	hyengnim	0.37	0.70	Used
146	과유불급	kwayupwulkup	0.37	-0.04	
147	전원주택	cenwencwuthayk	0.39	0.83	
148	할매	halmay	0.42	0.43	Used
149	영감탱이	yengkamthayngi	0.42	0.17	Used
150	골짜기	kolccaki	0.47	0.62	
151	브라자	pulaca	0.47	0.79	

152	이튿날	ithutnal	0.50	1.11	Used
153	빤스	ppansu	0.50	0.16	
154	흉악	hyungak	0.53	0.24	Used
155	마을	maul	0.53	0.62	Used
156	야단	yatan	0.53	0.68	Used
157	우여곡절	wuyekokcel	0.53	1.06	Used
158	따블	ttapul	0.56	0.65	
159	당신	tangsin	0.58	0.45	
160	시종일관	sicongilkwan	0.61	0.88	
161	회충약	hoychwungyak	0.63	0.40	Used
162	농민	nongmin	0.65	1.00	Used
163	풍년	phwungnyen	0.67	0.75	
164	후루츠	hwulwuchu	0.69	0.70	Used
165	뻰치	ppeynchi	0.69	0.82	
166	오라버니	olapeni	0.72	0.78	Used
167	보자기	pocaki	0.74	0.69	
168	총각	chongkak	0.74	0.65	
169	사진기	sacinki	0.74	0.27	
170	포대기	photayki	0.75	0.81	
171	고수부지	koswupwuci	0.80	1.17	
172	함마	hamma	0.82	-0.19	Used
173	쎄무	sseymwu	0.83	0.69	
174	주경야독	cwukyengyatok	0.83	1.12	
175	며느리	myenuli	0.84	1.33	Used
176	으뜸	uttum	0.84	0.44	Used
177	장사밑천	cangsamithchen	0.84	0.16	
178	나프킨	naphukhin	0.88	1.41	Used
179	냉이	nayngi	0.88	1.02	Used
180	앞잡이	aphcapi	0.88	0.81	Used
181	빠마	ppama	0.88	0.11	
182	장대비	cangtaypi	0.89	0.79	
183	테레비	theyleypi	0.89	0.58	
184	공구리	kongkwuli	0.90	0.16	
185	사라다	salata	0.93	0.51	

186	망년회	mangnyenhoy	0.94	1.13	Used
187	일품	ilphwum	0.94	0.91	Used
188	맷돌	maystol	0.94	1.05	Used
189	할망구	halmangkwu	0.94	0.53	Used
190	마누라	manwula	0.95	1.37	Used
191	대야	tayya	0.95	0.41	
192	동분서주	tongpwunsecwu	0.95	0.83	
193	캅셀	khapseyl	1.00	1.00	
194	들꽃	tulkkoch	1.00	0.76	
195	쥐약	cwiyak	1.00	0.21	
196	맛깔	maskkal	1.00	1.04	Used
197	흉년	hyungnyen	1.00	0.74	Used
198	팔자	phalca	1.00	0.90	
199	토사구팽	thosakwuphayng	1.00	0.35	
200	친손자	chinsonca	1.00	1.07	
201	추수	chwuswu	1.05	0.90	
202	달래	tallay	1.06	0.86	
203	비니루	pinilwu	1.06	0.75	
204	제부	ceypwu	1.06	0.92	
205	골무	kolmwu	1.07	0.72	
206	도라이버	tolaipe	1.07	0.49	
207	조카딸	cokhattal	1.11	1.11	
208	운수대통	wunswutaythong	1.12	1.20	Used
209	잡풀	capphwul	1.13	0.67	
210	갈퀴	kalkhwi	1.13	0.94	
211	후론트	hwulonthu	1.13	0.83	Used
212	가마솥	kamasoth	1.16	0.72	
213	미장원	micangwen	1.17	1.17	Used
214	약조	yakco	1.17	0.09	Used
215	미싱	mising	1.21	0.75	Used
216	란닝구	lanningkwu	1.21	0.10	Used
217	곡괭이	kokkwayngi	1.22	0.67	
218	장독대	cangtoktay	1.22	0.58	
219	오라이	olai	1.24	0.18	Used

220	빠다	ppata	1.25	0.06	
221	트랙터	thulaykthe	1.25	1.01	
222	자네	caney	1.26	0.53	
223	임자	imca	1.28	0.04	Used
224	명일	myengil	1.29	0.22	Used
225	매제	maycey	1.29	0.97	Used
226	이발소	ipalso	1.32	0.78	Used
227	저고리	cekoli	1.32	1.10	
228	모내기	monayki	1.32	0.54	Used
229	대청마루	taychengmalwu	1.32	0.83	
230	읍내	upnay	1.37	0.34	Used
231	뒷간	twiskan	1.37	0.11	
232	누이	nwui	1.37	0.82	Used
233	마후라	mahwula	1.38	1.12	Used
234	지게	cikey	1.39	0.97	
235	장터	cangthe	1.39	1.04	Used
236	안사람	ansalam	1.39	1.04	Used
237	언약	enyak	1.41	0.98	Used
238	쥐불놀이	cwipwulnoli	1.41	0.93	Used
239	방앗간	pangaskan	1.42	0.65	Used
240	사돈댁	satontayk	1.42	1.56	Used
241	얘야	yayya	1.44	1.11	Used
242	캬바레	khyapaley	1.44	0.75	Used
243	새참	saycham	1.44	0.88	Used
244	똥강아지	ttongkangaci	1.44	0.72	Used
245	도나쓰	tonassu	1.46	1.14	Used
246	당숙	tangswuk	1.47	0.57	Used
247	호롱불	holongpwul	1.47	0.49	Used
248	남가주	namkacwu	1.50	0.48	Used
249	구좌	kwucwa	1.50	1.76	
250	휴즈	hyucu	1.50	1.19	Used
251	작일	cakil	1.50	0.38	
252	여인숙	yeinswuk	1.50	0.98	Used
253	씨암탉	ssiamthalk	1.50	1.03	Used

254	빠께쓰	ppakkeyssu	1.50	0.59	Used
255	비료	pilyo	1.53	0.88	Used
256	조카사위	cokhasawi	1.56	1.28	Used
257	이보게	ipokey	1.58	0.01	Used
258	다방	tapang	1.58	0.26	Used
259	외손녀	oysonnye	1.58	1.49	Used
260	마을회관	maulhoykwan	1.58	0.61	Used
261	캇타	khastha	1.60	0.52	Used
262	뒤주	twicwu	1.60	0.55	Used
263	누이동생	nwuitongsayng	1.60	1.38	Used
264	호미	homi	1.61	1.30	Used
265	불란서	pwullanse	1.61	0.98	Used
266	헛간	heskan	1.63	0.83	Used
267	안사돈	ansaton	1.67	1.47	Used
268	여보시오	yeposio	1.67	0.44	Used
269	국민학교	kwukminhakkyo	1.67	1.07	Used
270	선산	sensan	1.69	1.47	Used
271	퇴비	thoypi	1.69	1.25	Used
272	샤쓰	syassu	1.70	1.23	Used
273	도마도	tomato	1.71	1.36	Used
274	절구	celkwu	1.72	0.74	Used
275	아범	apem	1.72	0.90	Used
276	복덕방	poktekpang	1.72	1.46	Used
277	곰방대	kompangtay	1.79	0.47	Used
278	영감쟁이	yengkamcayngi	1.79	1.14	Used
279	소쿠리	sokhwuli	1.79	1.06	Used
280	안내양	annayyang	1.81	0.75	Used
281	나성	naseng	1.83	0.43	Used
282	다라이	talai	1.86	0.79	Used
283	전축	cenchwuk	1.88	0.84	Used
284	새아가	sayaka	1.89	0.85	Used
285	쟁기	cayngki	1.89	0.59	Used
286	밤바	pampa	2.00	1.00	
287	뻬빠	ppeyppa	2.00	0.57	Used

288	어멈	emem	2.00	0.52	Used
289	이야기	iyaki	NA	NA	Used
290	마지막	macimak	NA	NA	Used
291	아저씨	acessi	NA	NA	Used
292	인터넷	intheneys	NA	NA	Used
293	분위기	pwunwiki	NA	NA	Used
294	아버지	apeci	NA	NA	Used
295	목소리	moksoli	NA	NA	Used
296	대부분	taypwupwun	NA	NA	Used
297	사람	salam	NA	NA	Used
298	얘기	yayki	NA	NA	Used
299	지금 2	cikum2	NA	NA	
300	생각	sayngkak	NA	NA	
301	여기	yeki	NA	NA	Used
302	시간	sikan	NA	NA	
303	친구	chinkwu	NA	NA	
304	문제	mwuncey	NA	NA	Used
305	그때	kuttay	NA	NA	
306	학교	hakkyo	NA	NA	Used
307	부분	pwupwun	NA	NA	
308	엄마	emma	NA	NA	Used
309	처음	cheum	NA	NA	Used
310	전화	cenhwa	NA	NA	
311	말씀	malssum	NA	NA	Used
312	누구	nwukwu	NA	NA	Used
313	필요	philyo	NA	NA	Used
314	소리	soli	NA	NA	Used
315	내용	nayyong	NA	NA	Used
316	대화	tayhwa	NA	NA	Used
317	시작	sicak	NA	NA	
318	방법	pangpep	NA	NA	Used
319	중요	cwungyo	NA	NA	
320	이름	ilum	NA	NA	Used
321	차이	chai	NA	NA	Used

322	마음	maum	NA	NA	Used
323	정보	cengpo	NA	NA	Used
324	표현	phyohyen	NA	NA	Used
325	설명	selmyeng	NA	NA	Used
326	과정	kwaceng	NA	NA	
327	만약	manyak	NA	NA	Used
328	자료	calyo	NA	NA	
329	학습	haksup	NA	NA	Used
330	나이	nai	NA	NA	Used
331	결혼	kyelhon	NA	NA	Used
332	운동	wuntong	NA	NA	Used
333	관계	kwankyey	NA	NA	
334	다섯	tases	NA	NA	
335	생활	saynghwal	NA	NA	
336	그림	kulim	NA	NA	
337	문화	mwunhwa	NA	NA	Used
338	중간	cwungkan	NA	NA	
339	상태	sangthay	NA	NA	Used
340	지금 1	cikum1	NA	NA	
341	회사	hoysa	NA	NA	
342	분석	pwunsek	NA	NA	
343	요즘	yocum	NA	NA	Used
344	아침	achim	NA	NA	Used
345	점수	cemswu	NA	NA	
346	이유	iyu	NA	NA	Used
347	의미	uymi	NA	NA	Used
348	방송	pangsong	NA	NA	Used
349	상황	sanghwang	NA	NA	Used
350	광고	kwangko	NA	NA	
351	저기	ceki	NA	NA	Used
352	하루	halwu	NA	NA	Used
353	아빠	appa	NA	NA	Used
354	얼굴	elkwul	NA	NA	Used
355	준비	cwunpi	NA	NA	Used

356	전체	cenchey	NA	NA	
357	교회	kyohoy	NA	NA	
358	자리	cali	NA	NA	Used
359	기분	kipwun	NA	NA	Used
360	여행	yehayng	NA	NA	Used
361	사랑	salang	NA	NA	
362	기본	kipon	NA	NA	Used
363	지역	ciyek	NA	NA	Used
364	외국	oykwuk	NA	NA	Used
365	장난	cangnan	NA	NA	
366	사진	sacin	NA	NA	Used
367	단체	tanchey	NA	NA	
368	정리	cengli	NA	NA	Used
369	느낌	nukkim	NA	NA	Used
370	관심	kwansim	NA	NA	Used
371	인간	inkan	NA	NA	Used
372	노래	nolay	NA	NA	Used
373	모습	mosup	NA	NA	Used
374	정신	cengsin	NA	NA	Used
375	주제	cwucey	NA	NA	
376	발달	paltal	NA	NA	Used
377	집단	ciptan	NA	NA	Used
378	결과	kyelkwa	NA	NA	Used
379	비교	pikyo	NA	NA	
380	반대	pantay	NA	NA	Used
381	특징	thukcing	NA	NA	
382	마찬가지	machankaci	NA	NA	Used
383	한국어	hankwuke	NA	NA	Used
384	상대방	sangtaypang	NA	NA	Used

APPENDIX II

Lexical stimuli in Experiment 4

Lexical stimuli used in Experiment 4 (96 pairs of target and competitor words) are listed below.

Phonemic transcriptions are based on "The Yale Romanization System".

Item #	Condition	Target word			Competitor word		
		Word	Phonemic transcription	Word age (ST score)	Word	Phonemic transcription	Word age (ST score)
1	Young target	꿀잼	kkwulcaym	-1.92	꿀차	kkwulcha	0.46
2	Young target	노잼	nocaym	-1.86	노파	nopha	1.22
3	Young target	득템	tuktheym	-1.78	득달	tuktal	0.60
4	Young target	페친	pheychin	-1.89	패착	phaychak	0.65
5	Young target	멘붕	meynpwung	-1.81	맨손	maynson	0.60
6	Young target	병맛	pyengmas	-1.97	병환	pyenghwan	1.36
7	Young target	밀당	miltang	-1.68	밀떡	milttek	0.83
8	Young target	고딩	koting	-1.51	고독	kotok	0.38
9	Young target	찐따	ccintta	-1.89	찐밤	ccinpam	0.70
10	Young target	베프	peyphu	-1.65	배필	payphil	1.57
11	Young target	먹튀	mekthwi	-1.89	먹성	mekseng	0.68
12	Young target	안습	ansup	-1.84	안달	antal	0.14
13	Old target	어멈	emem	1.97	어플	ephul	-1.54
14	Old target	쟁기	cayngki	1.38	쟁점	cayngcem	-0.05
15	Old target	절구	celkwu	1.32	절친	celchin	-1.14
16	Old target	샤쓰	syassu	1.17	샤방	syapang	-0.89
17	Old target	선산	sensan	1.40	선빵	senppang	-1.81
18	Old target	퇴비	thoypi	1.22	퇴짜	thoycca	-0.06
19	Old target	헛간	heskan	1.54	헛짓	hescis	0.49
20	Old target	호미	homi	1.32	호갱	hokayng	-1.84
21	Old target	뒤주	twicwu	1.80	뒤태	twithay	-0.84
22	Old target	당숙	tangswuk	1.54	당연	tangyen	0.00

23	Old target	언약	enyak	1.49	언니	enni	-0.16
24	Old target	잡풀	capphwul	1.27	잡솔	capsol	-1.56
25	Young comp	관계	kwankyey	-0.08	관종	kwancong	-1.92
26	Young comp	광고	kwangko	-0.22	광탈	kwangthal	-1.84
27	Young comp	노래	nolay	-0.08	노답	notap	-1.97
28	Young comp	단체	tanchey	-0.05	단톡	tanthok	-1.68
29	Young comp	대화	tayhwa	-0.08	대박	taypak	-0.86
30	Young comp	생각	sayngkak	0.00	생축	sayngchwuk	-1.49
31	Young comp	생활	saynghwal	0.00	생파	sayngpha	-1.78
32	Young comp	친구	chinkwu	-0.19	친추	chinchwu	-1.81
33	Young comp	시작	sicak	-0.14	시리	sili	-0.63
34	Young comp	얼굴	elkwul	-0.05	얼짱	elccang	-1.49
35	Young comp	엄마	emma	-0.38	엄빠	emppa	-1.84
36	Young comp	여행	yehayng	-0.08	여친	yechin	-1.27
37	Old comp	과정	kwaceng	0.00	과부	kwapwu	1.30
38	Old comp	나이	nai	0.00	나성	naseng	1.32
39	Old comp	누구	nwukwu	-0.08	누이	nwui	1.57
40	Old comp	다섯	tases	0.00	다방	tapang	1.46
41	Old comp	비교	pikyo	-0.03	비료	pilyo	0.95
42	Old comp	아빠	appa	-0.54	아범	apem	1.84
43	Old comp	얘기	yayki	-0.05	얘야	yayya	1.38
44	Old comp	자리	cali	-0.03	자네	caney	1.78
45	Old comp	장난	cangnan	-0.27	장터	cangthe	1.16
46	Old comp	전화	cenhwa	0.05	전보	cenpo	1.65
47	Old comp	지금	cikum	0.00	지게	cikey	1.30
48	Old comp	손짓	soncis	0.24	손주	soncwu	1.46
49	Filler	레알	leyal	-1.84	사랑	salang	-0.05
50	Filler	짤방	ccalpang	-1.89	요즘	yocum	-0.08
51	Filler	팀플	thimphul	-1.79	상황	sanghwang	0.00
52	Filler	토익	thoik	-1.74	소리	soli	0.00
53	Filler	겨털	kyethel	-1.68	외국	oykwuk	-0.16
54	Filler	라떼	lattey	-1.68	준비	cwunpi	-0.03
55	Filler	썩소	ssekso	-1.68	정리	cengli	0.08
56	Filler	훈남	hwunnam	-1.68	나무	namwu	0.00

57	Filler	강추	kangchwu	-1.63	만약	manyak	0.00
58	Filler	눈팅	nwunthing	-1.63	이유	iyu	0.00
59	Filler	폭망	phokmang	-1.60	내용	nayyong	0.00
60	Filler	놀토	noltho	-1.58	기본	kipon	-0.05
61	Filler	매제	maycey	1.29	지역	ciyek	0.05
62	Filler	뒷간	twiskan	1.37	사진	sacin	-0.11
63	Filler	읍내	upnay	1.37	필요	philyo	0.00
64	Filler	달래	tallay	1.06	처음	cheum	0.00
65	Filler	명일	myengil	1.29	점수	cemswu	-0.19
66	Filler	빠다	ppata	1.25	인간	inkan	0.14
67	Filler	미싱	mising	1.21	방송	pangsong	-0.05
68	Filler	약조	yakco	1.17	설명	selmyeng	-0.03
69	Filler	갈퀴	kalkhwi	1.13	회사	hoysa	-0.08
70	Filler	골무	kolmwu	1.07	이름	ilum	0.00
71	Filler	제부	ceypwu	1.06	중요	cwungyo	-0.03
72	Filler	추수	chwuswu	1.05	학습	haksup	-0.03
73	Filler	정보	cengpo	-0.08	치맥	chimayk	-1.47
74	Filler	방법	pangpep	0.00	열폭	yelphok	-1.79
75	Filler	정신	cengsin	0.03	열공	yelkong	-1.53
76	Filler	표현	phyohyen	-0.08	베플	peyphul	-1.84
77	Filler	시간	sikan	-0.03	극혐	kukhyem	-1.92
78	Filler	주제	cwucey	0.00	행쇼	hayngsyo	-1.92
79	Filler	문제	mwuncey	0.00	스펙	supheyk	-1.79
80	Filler	사람	salam	0.00	몸짱	momccang	-1.42
81	Filler	느낌	nukkim	-0.05	깜놀	kkamnol	-1.79
82	Filler	상태	sangthay	-0.05	안물	anmwul	-1.89
83	Filler	집단	ciptan	0.03	쌩얼	ssayngel	-1.42
84	Filler	그때	kuttay	0.03	썸녀	ssemnye	-1.89
85	Filler	교회	kyohoy	0.03	들꽃	tulkkoch	1.00
86	Filler	반대	pantay	0.00	맛깔	maskkal	1.00
87	Filler	의미	uymi	0.03	쥐약	cwiyak	1.00
88	Filler	중간	cwungkan	0.00	팔자	phalca	1.00
89	Filler	결혼	kyelhon	-0.05	흉년	hyungnyen	1.00
90	Filler	전체	cenchey	-0.05	아낙	anak	1.84

91	Filler	저기	ceki	0.00	오락	olak	0.41
92	Filler	여기	yeki	-0.08	화토	hwatho	0.74
93	Filler	분석	pwunsek	-0.30	냇가	nayska	0.95
94	Filler	특징	thukcing	-0.03	일꾼	ilkkwun	1.16
95	Filler	차이	chai	0.05	임자	imca	1.28
96	Filler	운동	wuntong	-0.11	만능	mannung	-0.27

APPENDIX III

Model structures and outputs in Experiment 4 (Models # 10–25)

- Model # 10 (data=early region, young and old competitor conditions)

glmer (Looks to comp ~ TalkerAge* WordAge + Time + TargetLocation +
(1+TalkerAge:WordAge+WordAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.100	0.310	-9.988	<.001
Talker age=young	-0.303	0.081	-3.722	<.001
Word age=young	0.280	0.223	1.257	.209
Time	10.532	0.340	30.989	<.001
Target location=right	1.335	0.080	16.774	<.001
Word age=young : Talker age=young	**0.080**	**0.249**	**0.321**	**.748**

- Model # 11 (data=late region, young and old competitor conditions)

glmer (Looks to comp ~ TalkerAge*WordAge + Time + TargetLocation +
(1+TalkerAge:WordAge+WordAge|Subject) + (1+TalkerAge:WordAge|Item))

	Estimate	Std. Error	z value	p value
(Intercept)	1.092	0.221	4.95	<.001
Word age=young	0.151	0.265	0.57	0.570
Talker age=young	-0.265	0.192	-1.38	0.169
Time	-9.367	0.258	-36.38	<.001
Target location=right	0.814	0.070	11.59	<.001
Word age=young : Talker age=young	**0.412**	**0.349**	**1.18**	**.238**

- Model # 12 (data=early region, young and old competitor conditions)

  glmer (Looks to target ~ TalkerAge* WordAge + Time + TargetLocation +
  (1+WordAge | Participant) + (1+TalkerAge | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.675	0.407	-9.021	<.001
Talker age=young	0.210	0.329	0.639	.523
Word age=young	0.285	0.247	1.153	.249
Time	11.740	0.390	30.114	<.001
Target location=right	-1.743	0.097	-17.940	<.001
Word age=young : Talker age=young	**-0.283**	**0.327**	**-0.864**	**.388**

- Model # 13 (data=late region, young and old competitor conditions)

  glmer (Looks to target ~ TalkerAge*WordAge + Time + TargetLocation +
  (1+TalkerAge:WordAge+WordAge | Subject) + (1+TalkerAge:WordAge|Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.481	0.232	-15.02	<.001
Word age=young	-0.068	0.213	-0.32	.750
Talker age=young	0.249	0.162	1.54	.124
Time	11.568	0.226	51.26	<.001
Target location=right	-0.372	0.055	-6.72	<.001
Word age=young : Talker age=young	**-0.364**	**0.261**	**-1.39**	**.164**

- Model # 14 (data=late region, young competitor condition)

  glmer (Looks to comp ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Participant) + (1+TalkerAge | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	1.190	0.376	3.163	.002
Talker age=young	**0.895**	**0.767**	**1.167**	**.243**
Time	-11.495	0.476	-24.170	<.001
Target location=right	0.527	0.116	4.565	<.001

- Model # 15 (data=late region, young competitor condition)

  glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Subject) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.730	0.316	-11.81	<.001
Talker age=young	**-0.496**	**0.360**	**-1.38**	**.169**
Time	13.173	0.382	34.51	<.001
Target location=right	-0.311	0.083	-3.73	<.001

- Model # 16 (data=early region, young competitor condition)

glmer (Looks to comp ~ TalkerAge + Time + TargetLocation +
(1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.854	0.378	-7.544	<.001
Talker age=young	**0.059**	**0.363**	**0.163**	**.870**
Time	10.123	0.424	23.896	<.001
Target location=right	0.834	0.099	8.407	<.001

- Model # 17 (data=early region, young competitor condition)

glmer (Looks to target ~ TalkerAge*Time + TargetLocation +
(1+TalkerAge | Subject) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.750	0.443	-8.461	<.001
Talker age=young	**-0.032**	**0.650**	**-0.049**	**.961**
Time	10.601	0.497	21.351	<.001
Target location=right	-1.759	0.135	-13.051	<.001
Talker age=young : Time	2.997	0.969	3.094	.002

- Model # 18 (data=early region, old competitor condition)

  glmer (Looks to comp ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-4.124	0.461	-8.953	<.001
Talker age=young	**0.186**	**0.521**	**0.358**	**.720**
Time	12.207	0.614	19.868	<.001
Target location=right	2.424	0.156	15.521	<.001

- Model # 19 (data=early region, old competitor condition)

  glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Subject) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-5.004	0.651	-7.692	<.001
Talker age=young	**0.272**	**0.714**	**0.381**	**.703**
Time	15.703	0.752	20.880	<.001
Target location=right	-2.286	0.161	-14.162	<.001

- Model # 20 (data=late region, old competitor condition)

  glmer (Looks to comp ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	1.142	0.247	4.624	<.001
Talker age=young	**-0.095**	**0.255**	**-0.371**	**.710**
Time	-8.584	0.316	-27.195	<.001
Target location=right	1.066	0.091	11.666	<.001

- Model # 21 (data=late region, old competitor condition)

  glmer (Looks to target ~ TalkerAge + Time + TargetLocation +
  (1+TalkerAge | Subject) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.502	0.287	-12.18	<.001
Talker age=young	**0.248**	**0.274**	**0.90**	**.366**
Time	10.948	0.292	37.55	<.001
Target location=right	-0.541	0.077	-7.06	<.001

- Model # 22 (data=early region, young and old target conditions)

  glmer (Looks to target ~ TalkerAge*CompStrength + Time + TargetLocation +
  (1 | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.970	0.351	-8.47	<.001
Talker age=young	0.602	0.107	5.64	<.001
Comp strength=weak	0.166	0.327	0.51	.611
Time	11.793	0.341	34.58	<.001
Target location=right	-1.631	0.080	-20.31	<.001
Talker age=young : Comp strength=weak	**0.021**	**0.146**	**0.15**	**.884**

- Model # 23 (data=early region, young and old target conditions)

  glmer (Looks to comp ~ TalkerAge*CompStrength + Time + TargetLocation +
  (1 | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-2.491	0.340	-7.33	<.001
Talker age=young	-0.185	0.094	-1.97	.048
Comp strength=weak	-0.143	0.361	-0.40	.692
Time	8.405	0.264	31.87	<.001
Target location=right	1.307	0.070	18.55	<.001
Talker age=young : Comp strength=weak	**-0.121**	**0.132**	**-0.91**	**.362**

- Model # 24 (data=early region, old target condition)

  glmer (Looks to target ~ TalkerAge*CompAge + Time + TargetLocation +
  (1 | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.860	0.485	-7.966	<.001
Talker age=young	1.028	0.120	8.566	<.001
Comp age=young	0.060	0.223	0.268	.789
Time	12.754	0.567	22.485	<.001
Target location=right	-2.509	0.144	-17.441	<.001
Talker age=young : Comp age=young	**-0.061**	**0.123**	**-0.493**	**.622**

- Model # 25 (data=early region, old target condition)

  glmer (Looks to comp ~ TalkerAge*CompAge + Time + TargetLocation +
  (1+CompAge | Participant) + (1 | Item))

	Estimate	Std. Error	z value	p value
(Intercept)	-3.574	0.470	-7.610	<.001
Talker age=young	-0.147	0.120	-1.223	.221
Comp age=young	-0.032	0.381	-0.084	.933
Time	10.350	0.492	21.055	<.001
Target location=right	1.382	0.123	11.242	<.001
Talker age=young : Comp age=young	**-0.154**	**0.122**	**-1.259**	**.208**

REFERENCES

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*(1), 544–552, doi:10.1121/1.1528172.

Allopenna, P. D., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439, doi:10.1006/jmla.1997.2558.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340–357, doi: 10.1037//0096-1523.10.3.340.

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, doi:10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48, doi:10.18637/jss.v067.i01.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133*, 2350–2366, doi:10.1121/1.4794366.

Blosfeld. M. E., & Bradley. D. C. (1981). *Visual and Auditory Word Recognition: Effects of Frequency and Syllahicity*. Paper presented at the Third Australian Language and Speech Conference, Melbourne, Austrailia.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *Quarterly Journal of Experimental Psychology, 65*(11), 2193–2220, doi: 10.1080/17470218.2012.693109.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software, 80*(1), doi:10.18637/jss.v080.i01.

Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Philadelphia: Benjamins, doi:10.1075/tsl.9.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. 65–85). Stanford, California: CSLI Publications, doi:10.1093/acprof:oso/9780195301571.003.0009.

Bybee, J. L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press, doi:10.1159/000085154.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change 14,* 261–90, doi: 10.1017/S0954394502143018.

Bybee, J. L., & Cacoullos, R. T. (2008). Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics, 1*, 399–413, doi:10.1515/shll-2008-1026.

Campbell-Kibler, K. (2007). Accent, (ING) and the social logic of listener perceptions. *American Speech, 82*(1), 32–64, doi:10.1215/00031283-2007-002.

Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics, 24*, 383–398, doi:10.1159/000262170.

Chambers, C., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and afforddances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30*(3), 687–696, doi: 10.1037/0278-7393.30.3.687.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268, doi:10.1016/j.wocn.2011.10.007.

Cho, T. (1996). *Vowel Correlates to Consonant Phonation: An Acoustic-perceptual Study of Korean Obstruents* (Unpublished doctoral dissertation). University of Texas Arlington, Arlington, TX.

Cho, T., Jun, S., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193–228, doi:10.1006/jpho.2001.0153.

Chodroff, E., & Wilsong, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47, doi:10.1016/j.wocn.2017.01.001.

Clopper, C. G., & Pisoni, D. B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*(1), 111–140, doi:10.1016/S0095-4470(03)00009-3.

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18*(2), 193–221, doi:10.1017/S0954394506060091.

Cohen-Goldberg, A. M. (2012). Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language, 67*(1), 184–198, doi:10.1016/j.jml.2012.03.007.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(1), 81–94, doi: 10.1037/0278-7393.19.1.81

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*(2), 633–664, doi:10.1016/j.cognition.2007.03.013.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2012). Word learning under adverse listening conditions: Context-specific recognition. *Language and Cognitive Processes, 27*(7–8), 1021–1038, doi:10.1080/01690965.2011.610597.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*(3), 264–285, doi:10.1016/j.jml.2011.06.005.

Cutler A., & Norris, D. (1979). Monitoring sentence comprehension. In W.E. Cooper & E. C. T. Walker (Eds.), *Sentence Processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

D’Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics, 19*(2), 241–256, doi:10.1111/josl.12115.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108*(3), doi:10.1016/j.cognition.2008.06.003.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317–367, doi: 10.1006/cogp.2001.0750.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507–534, doi: 10.1016/0010-0277(89)90013-9.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, Cognition, 30*(2), 498–513, doi: 10.1037/0278-7393.30.2.498.

Dahan, D., Tanenhaus, M. K., & Chambers, C. (2002). Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language 47(2), 292–314, doi: 10.1016/S0749-596X(02)00001-3.

Dart, S. N. (1987). An aerodynamic study of Korean stop consonants measurements and modeling. *The Journal of Acoustical Society of America, 81*(1), 138–147, doi:10.1121/1.395023.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech, 54*(1), 99–121, doi:10.1177/0023830910388017.

Drager, K., & Kirtley, M. J. (2016) Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. Babel (Eds.) *Awareness and Control in Sociolinguistic Research* (pp. 1–24), Cambridge: Cambridge University Press, doi:10.1017/CBO9781139680448.

Dufour, S., Brunelliere, A., & Frauenfelder, U. H. (2013). Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive Science 34*(3), 489–507, doi: 10.1111/cogs.12015.

Eckert, P. (1989), *Jocks and Burnouts: Social Categories and Identity in the High School*, New York: Teachers College Press.

Eckert, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Oxford, U.K.: Blackwell.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*(4), 453–476, doi:10.1111/j.1467-9841.2008.00374.x.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: languageand gender as community based practice. *Annual Review of Anthropology, 21*, 461–490, doi:10.1146/annurev.an.21.100192.002333.

Estes, W. K. (1993). Concepts, categories, and psychological science. *Psychological Science, 4*(3), 143–153, doi:10.1111/j.1467-9280.1993.tb00478.x.

Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics, 59*(6), 839–849, doi:10.3758/BF03205502.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409–438, doi:10.1016/j.wocn.2005.08.002.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Prychology: Human Perception and Performance, 10*(4), 526–540, doi: 10.1037/0096-1523.10.4.526.

Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science, 40*(3), 554–578, doi: 10.1111/cogs.12247.

Frazer, T. C. (1987). Attitudes toward regional pronunciation. *Journal of English Linguistics, 20*(1), 89–100.

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio and Electroacoustics, 16*(1), 73–77, doi:10.1109/TAU.1968.1161952.

Gahl, S. (2008). *Time* and *thyme* are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474–496, doi: 10.1353/lan.0.0035.

Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110–125, doi: 10.1037/0096-1523.6.1.110.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 144–158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 380–396, doi:10.1037/0096-1523.24.2.380.

Gelman, A., & Rubin., D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–511, doi:10.1214/ss%2F1177011136.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279, doi:10.1037/0033-295X.105.2.251.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267-283, doi: 10.3758/BF03204386.

Haiman, J. (1994). Ritualization and the development of language. In W. Pagliuca (Eds.), *Perspective on Grammaticalization* (pp. 3–28). Amsterdam: John Benjamins.

Han, J. (1996). *The Phonetics and Phonology of 'Tense' and 'Plain' Consonants in Korean* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*(1), 105–115, doi:10.1016/j.cogsci.2003.10.002.

Hay, J. (forthcoming). Sociophonetics: The role of words, the role of context, and the role of words in context. *Topics in Cognitive Science*.

Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology, 36*, 89–103, doi:10.1146/annurev.anthro.34.081804.120633.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865–892, doi:10.1515/LING.2010.027.

Hay, J. & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language, 92*(2), 298–330, doi: 10.1353/lan.2016.0036.

Hay, J., Nolan A., & Drager K. (2006a). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23*(3), 351–379, doi:10.1515/TLR.2006.014.

Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484, doi:10.1016/j.wocn.2005.10.001.

Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology, 5*, 1237. doi:10.3389/fpsyg.2014.01237.

Hombert, J.-M., Ohala, J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language, 55*(1), 37–58, doi:10.2307/412518.

Honjo, I., & Isshki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Archives of Otolaryngology, 106*(3), 149–150, doi:10.1001/archotol.1980.00790270013003.

Hooper, & Bybee, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change (pp. 96–105). In W. Christie (Eds.), *Current Progress in Historical Linguistics*. Amsterdam: North Holland.

Huettig, F, & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language 57*(4), 460–482, doi: 10.1016/j.jml.2007.02.001.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171, doi:10.1016/j.actpsy.2010.11.003.

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE, 9*(7), e100986, doi:10.1371/journal.pone.0100986.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*(6), 1939–1956, doi:10.1037/a0025641.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *The Journal of the Acoustical Society of America, 88*(2), 642–654, doi:10.1121/1.399767.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2005). Speaker normalization in speech perception, In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 145–165), Oxford, UK: Blackwell, doi: 10.1002/9780470757024.ch15.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499, doi: 10.1016/j.wocn.2005.08.004.

Johnson, K., & Beckman, M. E. (1996). Production and perception of individual speaking styles. *The Journal of the Acoustical Society of America, 100*(4), 2599, doi:10.1121/1.417601.

Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359–384, doi:10.1006/jpho.1999.0100.

Jun, S. (1993). *The Phonetics and Phonology of Korean Prosody* (Unpublished doctoral dissertation), The Ohio State University, OH.

Jun, S. (1996). The influence of the microprosody on the macroprosody: A case of phrase initial strengthening. *UCLA Working Papers in Phonetics, 92*, 97–116.

Jun, S. (2006). Phonological development of Korean: A case study. *UCLA Working Papers in Phonetics, 105*, 51–65.

Jun, S. (2011). Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 214–230). Somerville, MA: Cascadilla Press.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39–96). Cambridge, MA: MIT Press.

Kamide, Y., Altman, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133–156, doi:10.1016/S0749-596X(03)00023-8.

Kang, K., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America, 124*(6), 3909, doi:10.1121/1.2988292.

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, doi:10.1016/j.wocn.2014.03.005.

Kang, Y., & Han, S. (2013). Tonogenesis in early contemporary Seoul Korean: A longitudinal case study. *Lingua*, *134*, 62–74, doi: 10.1016/j.lingua.2013.06.002.

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition*, *139*, 28–49, doi:10.1016/j.cognition.2015.02.009.

Kim, M-D. (2004). Correlation between VOT and F0 in the perception of Korean stops and affricates. *Proceedings of Interspeech 2004 - ICSLP (8th International Conference on Spoken Language Processing)*, Jeju Island, Korea.

Kim, M-R. (2012). Tonogenesis in Korean: Some recent speculations on the sound change. *Korea Journal of Linguistics, 37*(2), 243–283, doi: 10.18855/lisoko.2012.37.2.001.

Kim, M-R. (2013). VOT merger and F0 maximization between the lax and aspirated stop in sound change. *The Linguistic Association of Korea Journal, 21*(2), 1–20.

Kim, M-R. (2014). Ongoing sound change in the stop system of Korean: A three- to two-way categorization. *Studies in Phonetics, Phonology and Morphology, 20*(1), 51–82, doi:10.17959/sppm.2014.20.1.51.

Kim, M-R., Beddor, P. C., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics, 30*(1), 77–100, doi:10.1006/jpho.2001.0152.

Kim, M-J. (2008). *The Phonetic and Phonological Development of Word-initial Korean Obstruents in Young Korean Children* (Unpublished doctoral dissertation), University of Washington.

Kim, S., Mitterer, H., & Cho, T. (2017). *An eye-tracking study on the role of prosody in phonological inferencing: A case of post- obstruent tensing rule in Korean*. Paper presented at the 25th Japanese/Korean Linguistics Conference, Honolulu, HI.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70,* 419–454, doi:10.2307/416481.

Kingston, J., Levy, J., Rysling, A., & Staub. A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception and Performance,* 42(12), 1969-1988: doi: 10.1037/xhp0000269.

Kleinschmidt, D. F. & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203, doi:10.1037/a0038695.

Kohler, K. J. (1982). F0 in the production of lenis and fortis plosives, *Phonetica, 39*(4–5), 199–218, doi:10.1159/000261663.

Kohler, K. J. (1984). Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica, 41*(3), 150–174. doi:10.1159/000261721.

Kohler, K. J. (1985). The perception of lenis and fortis plosives in French. A critical re-evaluation. *Phonetica, 42*(2–3), 116–123. doi:10.1159/000261742.

Kong, E., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics, 39*(2), 196–211, doi:10.1016/j.wocn.2011.02.002.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, 14, 93–101.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language, 62*(2), 83–97, doi:10.1016/j.jml.2009.10.001.

Labov, W. (1963). The social motivation of a sound change. *Word, 19*(3), 273–309, doi:10.1080/00437956.1963.11659799.

Labov, W. (1972). *Sociolinguistic Patterns*. Pittsburg: University of Pennsylvania Press.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*(2), 205–254, doi:10.1017/S0954394500000338.

Labov, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Malden, Oxford, UK: Blackwell.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America, 29*(1), 98–104, doi:10.1121/1.1908694.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*(3), 245–294, doi:10.1016/0010-0277(91)90008-R.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association, 42*(2), 145–169, doi:10.1121/1.3385294.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics, 41*(2), 117–132, doi:10.1016/j.wocn.2012.12.002.

Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society, 43*(5), 539–566, doi:10.1017/S0047404514000554.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–61, doi:10.1037/h0020279.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368, doi: 10.1037/h0044417.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, A. Marchal (Eds.), *Speech Production and Speech Modelling. NATO ASI Series (Series D: Behavioural and Social Sciences).* (Vol. 55, pp. 403–439). Dordrecht, Netherlands: Springer.

Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.

Luce, P A., & Pisoni, D. B. (1998) Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36, doi: 10.1097/00003446-199802000-00001.

Magnuson, J. S., Mirman, D., & Myers, E. (2013). Spoken word recognition. In D. Reisberg (Eds.), *The Oxford Handbook of Cognitive Psychology* (pp. 412–441). New York, USA: Oxford University Press, doi: 10.1093/oxfordhb/9780195376746.013.0027.

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition, 108*(3), 866–873, doi: 10.1016/j.cognition.2008.06.005.

Maran. L. R. (1973). On becoming a tone language: A Tibeto-Burman model of tonogenesis. In L. M. Hyman (Eds.), *Consonant Types and Tone, Southern California Occasional Papers in Linguistics.* (Vol. 1, pp. 98–114). Los Angeles, CA: Linguistics Program, University of Southern California.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25* (1–2), 71–102, doi: 10.1016/0010-0277(87)90005-9.

Marslen-Wilson, W. D. (1989). *Lexical representation and process*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Eds.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 148–172). Cambridge, MA: MIT Press.

Marslen-Wilson W. D., & Tyler L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1–71, doi: 10.1016/0010-0277(80)90015-3.

Marslen-Wilson W. D., Welsh A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29–63, doi: 10.1016/0010-0285(78)90018-X.

Marslen-Wilson, W. D., & Zwitserlood, R. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Prychology: Human Perception and Performance, 15*(3). 576–585, doi: 10.1037/0096-1523.15.3.576.

Martin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380, doi: 10.3758/BF03206780.

Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1996). Integration of multiple sources of information in language processing. In T. Inui, & J. L. McClelland (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 397–432). Cambridge, MA: MIT Press.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. M. Hyman (Eds.), *Consonant Types and Tone, Southern California Occasional Papers in Linguistics.* (Vol. 1, pp. 73–95). Los Angeles, CA: Linguistics Program, University of Southern California.

Mattys, S. L. (2013). Speech Perception. In D. Reisberg (Eds.), *The Oxford Handbook of Cognitive Psychology* (pp. 412-441). New York, USA: Oxford University Press, doi: 10.1093/oxfordhb/9780195376746.013.0026.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315, doi:10.1016/j.jml.2017.01.001.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–42doi: 10.1016/S0010-0277(02)00157-9.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 433–443, doi: 10.1037//0096-1523.17.2.433.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America, 85*(5), 2114–2134, doi:10.1121/1.2024119.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophys, 25*(6), 457–465, doi:10.3758/BF03213823.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Boca Raton, FL: Chapman and Hall/CRC.

Mueller, P. B., Sweeney, R. J., & Baribeau, L. J. (1984). Acoustic and morphologic study of the senescent voice. *Ear, Nose and Throat Journal*, 63(6), 292–295.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85*(1), 365–378, doi:10.1121/1.397688.

Munson, B., Jefferson, S. V., & McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America, 119*(1), 2427–2437, doi:10.1121/1.2173521.

Newman R. S., Clouse, S. A., Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America, 109*(3), 1181–1196, doi:10.1121/1.1348009.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass, 10*(11), 591–613. doi: 10.1111/lnc3.12207.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Psychology, 18*(1), 62–85, doi:10.1177/0261927X99018001005.

Nixon, J. S., Van Rij, J, Mok, P, Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103–125, doi: 10.1016/j.jml.2016.03.005.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234, doi:10.1016/0010-0277(94)90043-4.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357, doi:10.1037/0033-295X.113.2.327.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395, doi:10.1037/0033-295X.115.2.357.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370, doi: 10.1017/S0140525X00003241.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 700–708, doi:10.1037/0278-7393.14.4.700.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 3–27, doi:10.1037/0096-1523.17.1.3.

Nygaard, L. C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. *The Handbook of Speech Perception*. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*, Oxford, UK: Blackwell.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics, 39*(1), 59–67, doi:10.1016/j.wocn.2010.11.002.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Phonology: Learning, Memory, and Cognition, 19*(2), 309–328, doi:10.1037//0278-7393.19.2.309.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory of unfamiliar voices. *The Journal of the Acoustical Society of America 85*(2), 913-25, doi:10.1121/1.397564.

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research 4*(1), 10-29, doi:10.1044/jshr.0401.10.

Peterson, G. E., & Barney, H. I. (1952). Control methods used in a study of vowels. *The Journal of the Acoustical Society of America, 24*(2), 175–189.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency Effects and the Emergence of Linguistic Structure* (pp. 137–158). Amsterdam, Netherland: John Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII,* (pp. 101–139). Berlin: Mouton de Gruyter.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech, 46*(2–3), 115–154, doi:10.1177/00238309030460020501.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics 2*(1), 33–52, doi:10.1146/annurev-linguist-030514-125050.

Pinnow E., & Connine, C. M. (2014). Phonological variant recognition: Representations and rules. *Language and Speech*, *57*(1), 42–67, doi:10.1177/0023830913479105.

Pisoni, D. B. (1997). Some thoughts on “Normalization” in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp.9–32). San Diego, CA: Academic Press.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons with and across phonetic categories. *Perception and Psychophysics*, 15(2), 285–290, doi: 10.3758/BF03213946.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*(1), 19–36, doi:10.1016/j.jml.2009.02.005.

Podesva, R. J., Roberts, S. J., & Campbell-Kibler, K. (2002). Sharing resources and indexing meanings in the production of gay styles. In K. Campbell-Kibler, R. J. Podesva, S. J. Roberts, & A. Wong (Eds.) *Language and Sexuality: Contesting Meaning in Theory and Practice* (pp. 175–190). Stanford, California: CSLI Press.

Potter, R. K., & Steinberg, J. C. (1950). Toward the specification of speech. *The Journal of the Acoustical Society of America, 22*(6), 807–820, doi:10.1121/1.1906694.

Preston, D. R. (1993). Folk dialectology. In D. R. Preston (Eds.), *American Dialect Research* (pp. 333–378). Philadelphia: Benjamins.

Prinzmetal, W., & McCool, C. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General, 134*(1), 73–92, doi:10.1037/0096-3445.134.1.73.

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology, 70*, 1–30, doi:10.1016/j.cogpsych.2014.01.001.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81* (2), 275–280, doi:10.1037/h0027768.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance 23*(3), 651–66, doi:10.1037//0096-1523.23.3.651.

Rubin, P, Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perceprion and Pychophysics, 19*(5), 384–398.

Schertz, J., Kang, Y., & Han, S. (2016). *How much does the talker matter? Depends on who's listening: Age-related variability in the use of social information in speech perception*. Poster presented at The 15th Conference on Laboratory Phonology, Ithaca, NY.

Shin, J., & Cha, J. (2011). *Wulimal Soliuy Cheykyey (The Phonological System of Our Language)*. Seoul, Korea: Hankookmunhwasa.

Silva, D. J. (1992). *The phonetics and phonology of stop lenition in Korean* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Silva, D. J. (2002). Consonant aspiration in Korean: a retrospective. In S. Lee & G. K. Iverson (Eds.) *Pathways into Korean language and culture* (pp.447–469). Seoul: Pagijong Press.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology, 23*(2), 287–308, doi:10.1017/S0952675706000911.

Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America, 43*(1), 87–101, doi:10.1121/1.1910769.

Staum-Casasanto, L. (2008). *Does social information influence sentence processing?* Paper presented at the 30th Annual Meeting of the Cognitive Science Society, Washington, D.C.

Strand, E., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Eds.), *Natural Language Processing and Speech Technology* (pp. 14–26). Berlin, Germany: Mouton de Gruyter.

Sumner, M., Kim, S., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology, 4*, 1015, doi:10.3389/fpsyg.2013.01015.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*(3), 322–338, doi:10.1016/j.jml.2004.11.004.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501, doi:10.1016/j.jml.2009.01.001.

Swingley, D. (2005). Eleven-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*(5), 432–443, doi:10.1111/j.1467-7687.2005.00432.x.

Szakay, A., Babel, M., & King, J. (2016). Social categories are shared across bilinguals' lexicons. *Journal of Phonetics, 59,* 92–109, doi:10.1016/j.wocn.2016.09.005.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integratino of visual and linguistic information in spoken language comprehension. *Science*, 268 (5217), 1632–1634, doi: 10.1126/science.7777863.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580, doi:10.1023/A:1026464108329.

Tesink, C.M.J.Y., Petersson, K.M., Van Berkum, J., Van den Brink, D., Buitelaar, J., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience, 21*(11), 2085–2099, doi:10.1162/jocn.2008.21161.

Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324–333, doi: 10.1016/j.jcomdis.2009.03.001.

Tyler. L. K. (1984). The structure of the initial cohort: evidence from gating. *Perception & sychophysics, 36*(5), 417-427.

Van Berkum, J, Van den Brink, D., Tesink, C.M.J.Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591, doi:10.1162/jocn.2008.20054.

Van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *The Journal of Neuroscience, 32*(30), 10408–10412, doi:10.1523/JNEUROSCI.1337-12.2012.

Walker, A., & Hay J. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Laboratory Phonology, 2*(1), 219–237, doi:10.1515/labphon.2011.007.

Wang, X, Wang, J., & Malins, J. G. (2017). Do you hear ‘feather’ when listening to ‘rain’? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. *Cognition, 169*, 15–24, doi: 10.1016/j.cognition.2017.07.013.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167* (3917), 392–3, doi: 10.3758/BF03205070.

Weber A. & Scharenborg, O. (2012). Models of spoken-word recognition. *Cognitive Science, 3* (3), 387–401, doi: 10.1002/wcs.1178.

Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition, and Neuroscience, 30*(9), 1048–1060, doi: 10.1080/23273798.2014.946934.

Wells, J. (1982). *Accents of English*. Cambridge, England: Cambridge University Press.

Wright, J. (2007). *Laryngeal Contrast in Seoul Korean* (Unpublished doctoral dissertation). University of Pennsylvania.

Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, R. Temple, M. E. Beckman, & J. Kingston (Eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation*, (75–87). Cambridge UK: Cambridge University Press.

Yaegor-Dror, M., & Kemp, W. (1992). Lexical classes in Montreal French. *Language and Speech, 35*(3), 251–293, doi:10.1177/002383099203500301.

Yi, S. (2015). *Sociophonetic Variations in Korean Constituent Final –ko and –to* (Unpublished doctoral dissertation). University of Hawai‘i at Mānoa.

Zhuang, J., Randall, B., Stamatakis, E. A., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The interaction of lexical semantics and cohort competition in spoken word recognition: An fMRI study. *Journal of Cognitive Neuroscience, 23*(12), 3778–3790, doi: 10.1162/jocn_a_00046.

Zwitserlood, P. (1985). *Activation of word candidates during spoken word-recognition*. Paper presented at Psychonomic Society Meetings, Boston, MA.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context on spoken-word processing. *Cognition, 32*(1), 25–64, doi: 10.1016/0010-0277(89)90013-9.