



Sociophonetic realizations guide subsequent lexical access

Jonny Kim, Katie Drager

University of Hawai‘i at Mānoa, USA

jonnykim@hawaii.edu, kdrager@hawaii.edu

Abstract

Previous studies on spoken word recognition suggest that lexical access is facilitated when social information attributed to the voice is congruent with the social characteristics associated with the word. This paper builds on this work, presenting results from a lexical decision task in which target words associated with different age groups were preceded by sociophonetic primes. No age-related phonetic cues were provided within the target words; instead, the non-related prime words contained a sociophonetic variable involved in ongoing change. We found that age-associated words are recognized faster when preceded by an age-congruent phonetic variant in the prime word. The results demonstrate that lexical access is influenced by sociophonetic variation, a result which we argue arises from experience-based probabilities of covariation between sounds and words.

Index Terms: lexical access, sociophonetics, sound change

1. Introduction

Through everyday interactions, cognitive associations between social, phonetic, and lexical information are formed and shaped by experience. Listeners can then exploit these associations during speech perception. For example, when exposed to an ambiguous auditory signal, listeners' segment-level categorization can be influenced by social factors such as the age [1] and gender [2, 3] attributed to the talker. Such contextual effects also seem to play a role in lexical access. Results from short-term repetition priming tests suggest that facilitated recognition occurs when the incoming acoustic signal resembles the existing mental representation of individual listeners [4, 5]. For example, listeners' lexical access depends on a combination of whether /r/ is realized in the signal and the listeners' prior exposure to /r/-less varieties [4]. Likewise, there is evidence that English words are recognized more quickly when the age of the people who tend to produce a word and the age of the talker are congruent [6], a result that has been replicated for Korean [7]. Taken together, the results indicate that listeners' expectations about talkers influence lexical access.

In previous experiments that found an effect of talker age on lexical access [6, 7], words were produced by talkers of different ages, meaning that there were multiple phonetic cues indexing the talkers' ages. This raises a number of questions: First, are listeners' expectations guided by the phonetic realizations of sounds involved in a change-in-progress when that variation is the only cue to the talker's age? Second, does the variation need to be inherent in the target word itself, or can a previously occurring sociophonetic prime affect subsequent processing of a target word? And, finally, do effects from such priming extend to same-talker trials in which no cues to talker age are found? In this paper, we report results

from a lexical decision task that investigates whether recognition of age-associated words in Korean is influenced by realizations of a variable involved in ongoing sound change, while controlling talker age.

Apparent-time research on the contemporary Seoul Dialect reports that the traditional VOT-based distinction between phrase-initial lax (/p, t, k/) and aspirated (/p^h, t^h, k^h) stops is gradually being replaced by an F0 distinction, which was previously a secondary feature [8, 9, 10, 11]. There is no apparent change in the F0/VOT dimension of lax stops, but aspirated stops are going through a gradual decrease in VOT and an increase in F0 at voice onset, so that younger speakers tend to have a shorter VOT and a higher pitch in aspirated stops compared to older speakers. Additionally, F0 is raised in the innovative realizations of tense stops (/p*, t*, k*/). An overview of these changes is shown in Table 1. These changes are led by younger female speakers [12] and are spreading into other dialects so that segment perception in even the conservative dialect regions is affected [13].

Table 1: Overview of VOT and F0 found in younger and older speaker's realizations of aspirated and tense stops.

	Prime	VOT	Pitch
ASP	[p ^h alaŋ] 'blue'	old > young	old < young
TENSE	[p*alkaŋ] 'red'	no difference	old < young
LAX	[pamsek] 'brown'	no difference	no difference

Because the change is ongoing, older and younger speakers of the Seoul Dialect produce different realizations of aspirated and tense stops. In the lexical decision task reported in this paper, we manipulated both F0 and VOT in aspirated and tense plosives to determine whether these cues – in the absence of other age-associated phonetic cues – would affect lexical access. We predict that recognition of age-associated target words is facilitated when preceded by an age-congruent stop variant.

2. Method

2.1. Participants

40 Korean-native college students (27 females and 13 males, age: 18-27) took part in the experiment. All but five participants listed Seoul Dialect as their most frequently used dialect, and these five participants (3 males and 2 females) listed Kyeongsang Dialect as primary and Seoul Dialect as secondary. Data were collected from two locations: 20 at Chung-Ang University in Seoul, Korea and 20 at the

University of Hawai'i at Mānoa. All Hawai'i participants had lived in Hawai'i for less than two years.

2.2. Procedure

Each trial was composed of two tasks in which words produced by a single talker were played. The first task was an identification task in which participants indicated which color they heard. The second task was a lexical decision task. For the identification task, one of the three color terms, *red*, *blue*, or *brown*, ([p*alkan], [p*alaŋ], and [pamsek], respectively) was played, and participants were instructed to press the button in the corresponding color (marked on the button box) with their recessive hand. Following the button press, a real word or non-word target was played, and then participants pressed "Real" or "False" with their dominant hand, indicating whether what they heard was a real word. Participants were instructed to press each button only with a designated finger; for the lexical decision responses, participants pressed the "Real" button with the index finger and "False" button with the thumb of their dominant hand. Blocks were made up of 135 trials each, and participants were able to choose to take a short break between blocks. Except for the break, participants were instructed not to move their hands off the response box.

2.3. Auditory stimuli

The auditory stimuli presented to each participant were composed of three priming stimuli and 540 target stimuli, which were produced by two female critical talkers (age: 40 and 41) and two male distracter talkers (age: 34 and 36).

2.3.1. Priming stimuli

The talkers were instructed to produce the color terms in an H+L% prosodic structure, and no instructions about VOTs were given. Recordings of the color terms produced by the critical talkers were manipulated to create four priming stimuli (two color terms in an old guise and a young guise). Using Praat, periodic zero-crossing cycles were selected 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. For F0 manipulation, pitch contours were manipulated using the "To Manipulation..." function in Praat.

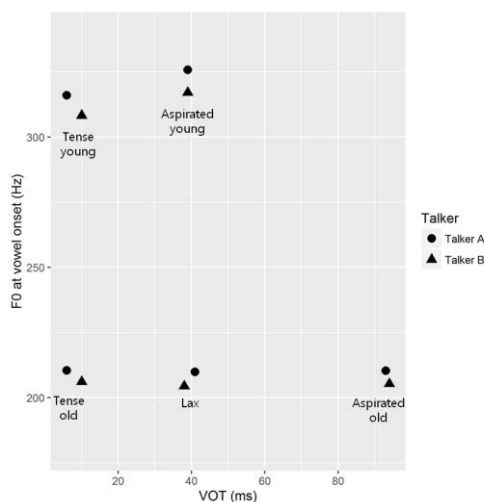


Figure 1: VOT/F0 distribution of priming stimuli.

The manipulated tokens of color terms are plotted in a VOT/F0 dimension in Figure 1. For the young guise, VOT of the aspirated stop (39ms, 39ms) was comparable with that of the lax stop (41ms, 38ms), mimicking the VOT merger of younger speakers, whereas the aspirated stop in the old guise had a longer VOT (93ms, 94ms). The young guise also had a higher pitch at the voice onset for tense (316Hz, 308Hz) and aspirated tokens (326Hz, 317Hz) than the old guise tokens (tense: 210Hz, 206Hz, aspirated: 210Hz, 205Hz). The VOT/F0 values used in the stimuli are within a normal range of production data reported in the literature [9, 11].

Unvarying lax tokens were also played in order to provide the listeners with a reference point in terms of VOT and F0 for each talker. Including the lax tokens also allowed us to test whether exposure to the sociophonetic variable in aspirated and tense trials would activate talker age and then be generalized to the lax trials produced by that talker. Because participants were only exposed to one guise per talker, such a generalization is conceivable.

2.3.2. Target stimuli

Target stimuli were composed of 270 real words and 270 non-words. The words were selected from those in [7], in which real word items were assigned a *word age*, a continuous variable indicating whether young Koreans believed the word is used by older or younger people. Lower values (from -2 to -1) indicate young-associated words, whereas higher values (from 1 to 2) indicate old-associated words. Fillers were chosen from the most frequent words in the Sejong spoken corpus.

Table 2: Summary of target stimuli.

Lexical status	Talker type	Word age	Word onset
Real words (N=270)	Critical (N=146)	37 old, 39 neutral, 34 young, 36 fillers	Non-obstruent
	Distract (N=124)	31 old, 31 neutral, 31 young, 31 fillers	Obstruent
Non-words (N=270)	Critical (N=124)	NA	Non-obstruent
	Distract (N=146)	NA	Obstruent

As shown in Table 2, 270 items with a non-obstruent onset (146 real words and 124 non-words) were produced by the critical talkers. The other 270 items (124 real words and 146 non-words) were ones that do contain an initial obstruent and were produced by two male distracter talkers.

2.4. Design

All items were counterbalanced by talker ID (Talker A and B) and guise (old and young) into four lists in a Latin Square design. All items appeared in a random order for each participant. The three color terms were evenly paired with the 540 items and were not counterbalanced across items.

3. Results

Reaction times (RTs) were obtained from the onset of target stimuli to the button press for real word trials with correct responses to both tasks (N=10,645). Trials with an inaccurate color press were excluded because preliminary analysis revealed that target word recognition was significantly slower when listeners pressed an incorrect color button, an effect we attribute to a difference in attention. Data points with reaction times below 2/3 of the target duration and over 5,000ms were removed (N=18). Based on the remaining tokens, responses over three standard deviations from the mean by participant were removed (N=165). Finally, responses from distracter items were removed, resulting in 4,238 tokens remaining for analysis.

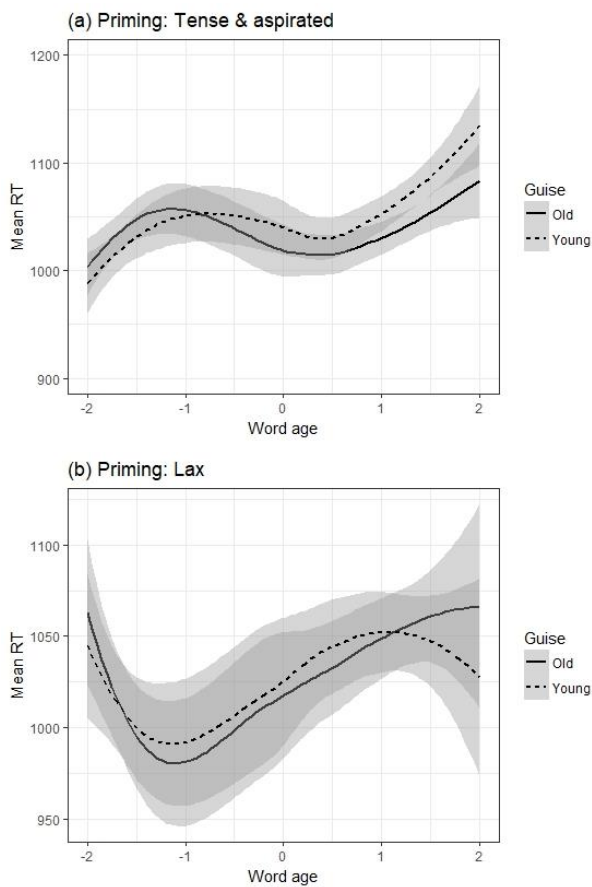


Figure 2: Mean RTs by word age and guise for words preceded by (a) tense or aspirated priming and (b) lax priming. Covariates are estimated using the loess smooth method. Grey areas indicate standard errors.

In Figure 2, means of raw RTs from (a) trials with primes containing tense and aspirated stops (N=2,845) and (b) trials with lax primes tend to be longer as word age increases. This is likely due to the young age of the participants, who have minimal exposure to words with older word ages. One exception is the set of youngest words preceded by a lax prime; further work is needed to interpret this subsection of the results. Also evident in Figure 2(a) is a cross-over interaction between word age and guise; responses for words that are highly associated with younger people (word age lower than -

1) were faster when preceded by the young guise, while words with higher word age were recognized faster when preceded by the old guise. In contrast, such an interaction is not found for trials with primes containing the (unmanipulated) lax stop (N=1,393), as shown in Figure 2(b).

To test the significance of the effects observed above, a linear mixed effects model was fit in R using the lme4 package [14]. Treated as the dependent variable were raw RTs for tokens with primes containing the tense and aspirated stops. Fixed effects tested in the model include talker ID, trial order, trial block, test location, computer station, participant sex, participant handedness, participant dialect, and item list. All two-way and three-way interactions between factors were tested, and model fit was tested using ANOVAs. Only interactions and factors that improved the model's fit were included in the final model. In the final model, guise (binary, deviation coded) and word age were included as test variables, both as main effects and in interaction. Other fixed effects included in the final model were plosive ID (tense or aspirated), word duration (as a control variable), an interaction between test location (binary) and word age (continuous).

For the random effects structure, we began from the maximal structure justified by the design [15], but by-item slope for guise was excluded, as the model failed to converge. The random effects included in the final model were (1) by-participant slopes for guise, word age, and their interaction, (2) by-item slopes for an interaction between guise and word age, and (3) intercepts for by-participant and by-item variance.

Table 3. Summary of results of mixed effects model.

	Estimate	S.E.	t-val	p-val
(Intercept)	770.112	43.744	17.605	<.001
Plosive ID=tense	-7.215	19.794	-0.364	.715
Word duration	0.449	0.062	7.215	<.001
Guise=young	5.355	3.872	1.383	.167
Word age	15.989	8.052	1.986	.047
Test location=Seoul	-16.423	5.855	-2.805	.005
: Word age				
Guise=young	6.937	2.922	2.374	.018
: Word age				

As shown in Table 3, RTs were not significantly different between trials with the tense prime and with the aspirated prime ($p=.715$). Word duration had a significant main effect, and is included as a control variable ($p<.001$). Guise did not influence RTs significantly ($p=.167$), but a significant main effect of word age indicates that words were recognized more slowly, as word age increased ($p=.047$). Surprisingly, there is a significant interaction between test location and word age, indicating that participants in Seoul were faster than those in Hawai'i when responding to words with higher word age ($p=.005$). When a main effect of place was added in the model, its effect was not significant ($\beta=-31.728$, $s.e.=21.270$, $p=.136$), indicating that facilitated responses occurred mainly on words with higher word age, rather than as a general processing advantage for participants in Seoul.

Importantly, there was a significant interaction between guise and word age; when primed with the young-associated phonetic variant, participants took longer to recognize words with higher word ages ($p=.018$), confirming our main hypothesis. We find no evidence, however, to support our secondary hypothesis; in a separate model fit to trials preceded

by the unvarying lax primes, an interaction between guise and word age was not significant ($\beta=-1.955$, $SE=7.728$, $p=.800$).

4. Discussion

Our results are consistent with previous work demonstrating that lexical access is influenced by phonetic primes when the lexical item and the phonetic variant index similar social information. Further, the results provide evidence that such an effect can arise even when the prime is restricted to a single sociophonetic variable and when it occurs prior to the test item. These results highlight the role that socio-indexical phonetic variation plays in the processing of lexical information.

There are three possible avenues through which we may have observed the priming effect: (1) upon encountering socially-conditioned phonetic realizations, abstract representations of associated social information are activated, and the activation then spreads to associated lexical items, increasing the speed at which these items can be identified as real words (2) phonetically-rich multi-word representations may be stored and accessed, or (3) sociophonetic variants may directly index lexical items that tend to be produced by the same speakers as those who produce the socially-associated realizations.

Based on (1), we made a secondary hypothesis that listeners would generalize the talker's perceived age across tokens produced by that talker and that, as a result, listeners would exhibit an effect of priming condition on tokens that were not immediately preceded by an age-associated variant. However, we find no evidence of this in our data. One possible explanation of this lack of an effect is that the listeners may not have been aware that they were hearing only two female talkers, decreasing the likelihood that social information would be generalized to the lax trials. However, it is also possible that abstract social information about talker age did not contribute to the observed effect, making either of the latter two interpretations (2) or (3) more likely.

In the second interpretation, phonetically-rich multi-word chunks of speech are stored and activated, resulting in a direct link between the target word and phonetic detail in the preceding word. However, we find this possibility to be unlikely, especially due to the strange combination of the prime + target pairs (e.g., *red* + *elementary school*), multi-word utterances that were unlikely to have been encountered previously.

The third interpretation, which we find to be the most compelling, suggests a direct link between phonetic realizations and lexical items, even when the word itself does not contain the phonetic information. In this account, encountering particular phonetic realizations boosts the probability of encountering lexical items that are used by the same individuals who tend to produce the phonetic realizations in question. Such a relationship could arise in usage-based models in which lexical access is guided by contextual probabilities (e.g., probabilities that arise from covariance between phonetic realizations and lexical items). Thus, the results do not necessarily require that mental representations of the associated social information be activated, even though the phonetic detail is not found in the target word itself.

Along with previous work [4,6,7], the results also have methodological implications. If sociophonetic variation influences lexical access, sociophonetic realizations should be

considered in experiments investigating lexical access, even when the sociophonetic variant occurs in the prime word rather than the target word and when the prime is not semantically-related to the target.

In addition, it appears that test location also influenced recognition of old-associated words. Slower recognition of old-associated words by participants in Hawai'i could be due to lack of recent exposure; however, this is unlikely since all but two participants left Korea less than a month before taking part in the experiment. Rather, we attribute it to an effect of location itself. Being tested in a foreign country could lead to difference in activation level or prior expectation for the probability of encountering an old-associated word.

5. Conclusions

This paper demonstrates that lexical access is guided by phonetic variation, and that an effect can arise even when there is only a single sociophonetic variable manipulated across conditions and when the phonetic cues are not found in the target words themselves. We interpret these findings as evidence that listeners are sensitive to covariance of sociophonetic variants and lexical items, and we argue that the results highlight the role of experience-based probabilistic information in lexical access.

6. Acknowledgements

This study was funded by the Bilinski Dissertation Fellowship. We also would like to thank the Department of English Language and Literature at Chung-Ang University and the Center for Asia-Pacific Exchange at the University of Hawai'i at Mānoa for providing access to participant pools.

7. References

- [1] K. Drager, "Speaker age and vowel perception," *Language and Speech*, vol. 54, no. 1, pp. 99–121, 2011.
- [2] E. Strand and K. Johnson, "Gradient and visual speaker normalization in the perception of fricatives," *Natural Language Processing and Speech Technology*, ed. D. Gibbon, pp. 14–26. Berlin, Germany: Mouton de Gruyter, 1996.
- [3] K. Johnson, E. Strand, and M. D'Imperio, "Auditory-visual integration of talker gender in vowel perception," *Journal of Phonetics*, vol. 27, pp.359–384, 1999.
- [4] M. Sumner and A. G. Samuel, "The role of experience in the processing of cross-dialectal variation," *Journal of memory and language*, vol. 60, pp. 487–501, 2009.
- [5] A. Szakay, M. Babel, and J. King, "Social categories are shared across bilinguals' lexicons," *Journal of Phonetics*, vol. 59, pp. 92–109, 2016.
- [6] A. Walker and J. Hay, "Congruence between 'word age' and 'voice age' facilitates lexical access," *Laboratory Phonology*, vol. 2, no. 1, pp. 219–237, 2011.
- [7] J. Kim, "Perceptual associations between words and speaker age," *Laboratory Phonology*, vol. 7, no. 1, pp. 18, 2016.
- [8] D. J. Silva, "Acoustic evidence for the emergence of tonal contrast in contemporary Korean," *Phonology* vol. 23, pp. 287–308, 2006.
- [9] K. H. Kang and S. G. Guion, "Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies," *The Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 3909–3917, 2008.
- [10] Y. Kang and S. Han, "Tonogenesis in early contemporary Seoul Korean: A longitudinal case study" *Lingua*, vol: 134, pp.62–74, 2013.

- [11] Y. Kang, "Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study" *Journal of Phonetics*, vol: 45, pp. 76–90, 2014.
- [12] E. Oh, "Effects of speaker gender on voice onset time in Korean stops" *Journal of Phonetics*, vol: 39, no. 1, pp. 59–67, 2011.
- [13] J. Schertz, Y. Kang, and S. Han, "How much does the talker matter? Depends on who's listening: Age-related variability in the use of social information in speech perception," in *The 15th Conference on Laboratory Phonology, Ithaca, NY, poster presentation*, 2016.
- [14] D. Bates, M. Maechler, B. Bolker and S. Walker, "lme4: Linear mixed-effects models using Eigen and S4". R package version 1.1-12. <http://CRAN.R-project.org/package=lme4>
- [15] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal" *Journal of Memory and Language*, vol: 68, pp. 255–278, 2013.