Intoxication and pitch control in tonal and non-tonal language speakers

Kevin Tang,[1, a)] **Charles B. Chang,**[2] **Sam Green,**[3] **Kai Xin Bao,**[4] **Michael Hindley,**[3] **Young Shin Kim,**[5] **and Andrew Nevins**[3]

1) *Department of English Language and Linguistics, Institute of English and American Studies, Heinrich-Heine-University, Düsseldorf, 40225, Germany*

2) *Department of Linguistics, Boston University, Boston, MA 02215, USA*

3) *Department of Linguistics, University College London, London WC1N 1PF, UK*

4) *Department of Experimental Psychology, University College London, London WC1N 1PF, UK*

5) *Language and Culture Laboratory, Cheongju, 28775, Republic of Korea*

kevin.tang@hhu.de,
cc@bu.edu,
sam.green.10@ucl.ac.uk,
kbao6211@gmail.com,
mike@mikehindley.co.uk,
young.s.kim@alumni.ucl.ac.uk,
a.nevins@ucl.ac.uk

(Dated: 6 May 2022)

Abstract: Alcohol intoxication is known to affect pitch variability in non-tonal languages. In this study, intoxication's effects on pitch were examined in tonal and non-tonal language speakers, in both their native language (L1; German, Korean, Mandarin) and nonnative language (L2; English). Intoxication significantly increased pitch variability in the German group (in L1 and L2), but not in the Korean or Mandarin groups (in L1 or L2), although there were individual differences. These results support the view that pitch control is related to the functional load of pitch and is an aspect of speech production that can be advantageously transferred across languages.

---

[a] Author to whom correspondence should be addressed.

1. Introduction

Consumption of alcohol, a central nervous system depressant, has long been known to affect speech production, including aspects of vocal pitch.[1–3] For example, research on English speakers in a repetition task found that intoxication up to a blood alcohol concentration (BAC) of at least 0.10%, while having little effect on overall mean fundamental frequency (f_0) level for most speakers, consistently led to higher f_0 variability than in sober (i.e., unintoxicated) speech, an effect "suggesting less precise control of the rate of vocal cord vibration" under intoxication.[4] Although it is not clear whether changes in f_0 variability can be used to reliably identify intoxication, the finding of increased f_0 variability in intoxicated speech has been replicated in other studies of English, which have also evinced individual differences in the presence and/or directionality of an intoxication effect on mean f_0 level.[5,6]

Previous studies of other languages have contributed to a fuller picture of how speakers' pitch control may be affected by intoxication, suggesting that there may be considerable crosslinguistic variability in this regard. On the one hand, German speakers have mostly shown an increase in mean f_0 and f_0 range under intoxication, but also a number of individual differences.[7] On the other hand, Japanese speakers have shown a significant decrease in mean f_0, as well as a non-significant tendency toward an expanded f_0 range.[8] One potential contributor to such crosslinguistic variability is typological variation among languages in the functional role played by a given cue. In the case of f_0, this may serve primarily to signal pragmatic distinctions at the sentence/utterance level (INTONATION languages; e.g., English), lexical contrasts in part of the vocabulary (PITCH ACCENT languages; e.g., Japanese), or lexical contrasts across the entire vocabulary (TONAL languages; e.g., Mandarin Chinese). The fact remains, however, that there are very few acoustic studies of intoxicated speech in languages that are not English, thus limiting any typological account of crosslinguistic variability.

Apart from typological differences in the role of f_0, another potential contributor to variation in the effects of intoxication is experience (and proficiency) in the target language. In particular, it has been suggested that effects of intoxication differ for one's native language (L1; generally an early-learned and relatively strong language) and a nonnative language (L2; generally a later-learned and relatively weaker language). For instance, whereas intoxication has generally been found to negatively affect production in a speaker's L1, it was found to positively affect production in an unfamiliar L2 (as measured by global accent ratings), which was attributed to intoxication modifying a speaker's "language ego" in a manner facilitating authentic (i.e., native-like) L2 pronunciation.[9] Along the same lines, Dutch speakers have shown a detrimental effect of alcohol consumption on the clarity of their L1 (Dutch) speech but no such effect on the perceived nativelikeness of their L2 (English) speech.[10]

Notably, the L1-L2 disparities in intoxication effects at a global level stand in contrast to the findings of acoustic studies of bilingual speech, which often provide evidence of similarities—and, by implication, interconnections—between the L1 and L2, including in aspects of prosody.[11–14] Findings showing crosslinguistic influence related to pitch control have been reported for f_0 alignment in L1 Dutch-L2 Greek and L1 German-L2 English speakers,[15,16] f_0 range for L1 Welsh-L2 English speakers (albeit mostly for males),[17] and

f_0 level for L1 English-L2 Korean speakers,[18–20] consistent with the view that there is a crosslinguistically "shared control mechanism for f_0 modulation."[18] Few studies, however, have examined f_0 variability crosslinguistically, much less in conditions that undermine articulatory control such as intoxication.

Thus, in the current study we bring together typological and acquisition-related concerns to ask two questions regarding the effects of alcohol intoxication on speech production. First, does intoxication affect pitch variability similarly across languages that differ in the level of pitch control they require, such as tonal and non-tonal languages (**Q1**)? Second, do sequential bilinguals of diverse L1-L2 backgrounds show similar effects of intoxication on pitch variability in their L1 and L2 (**Q2**)? To investigate these questions, we carried out a bilingual acoustic study of intoxicated speech produced by L2 English speakers from three L1 backgrounds: German (an intonation language), Korean (an intonation language with tone-like contrasts in certain phrase-prosodic positions),[21,22] and Mandarin (a tonal language). Under the assumption that speakers' articulatory control of a phonetic cue reflects the cue's relative functional load in the language (i.e., the unique linguistic burden it bears in signaling contrasts),[23] L1 Mandarin speakers will be predisposed toward greater pitch control than L1 German or L1 Korean speakers, because the relative functional load of pitch is the highest in Mandarin.[24] This leads to the hypothesis that intoxication will impact the variability of f_0 (the acoustic correlate of pitch) less for L1 Mandarin speakers than for L1 German or Korean speakers (**H1**). Furthermore, if f_0 is indeed modulated at least in part by a control mechanism that is shared between languages, this leads to the hypothesis that, for all L1 groups, effects of intoxication on f_0 variability will look similar in the L1 and L2 (**H2**).

2. Methods

2.1 Participants

In order to be included in the study, participants had to: (a) identify as a native speaker of one of the target L1s, (b) identify as an L2 speaker of English, (c) be at least 21 years old, (d) not have been diagnosed with hearing deficits or speaking disorders, (e) not be currently pregnant, and (f) not be struggling with alcohol-related problems of any kind (e.g., alcoholism). The three L1 groups comprised native speakers of German ($N = 8$; 4f, 4m; $M_{age} = 27.1$ yr, $SD = 4.3$), Korean ($N = 8$; 8f, 0m; $M_{age} = 27.1$ yr, $SD = 3.8$), and Mandarin ($N = 17$; 10f, 7m; $M_{age} = 23.8$ yr, $SD = 1.5$) who were born and raised/educated in an L1-dominant environment (i.e., Germany, South Korea, mainland China, respectively) and self-reported their L2 English level as fluent. In the Korean group, most participants (7) were from Seoul or the surrounding Gyeonggi province, with one from the North Gyeongsang province; thus, most spoke Seoul Korean or a similar dialect. In all groups, most participants were students who had been living in the UK for 1–2 years at the time of the study.

Two types of objective data on participants' L2 English proficiency were collected. First, International English Language Testing System (IELTS) scores were collected if available. IELTS scores were high overall and did not differ significantly between groups (Welch-corrected two-sample $|t|\text{s} < 1.7$, $p\text{s} > 0.05$). The group means were all in the 7.0 band of the IELTS scale, which indicates being a "good" user of the English language and translates to a “lower advanced” (C1) level of proficiency in the Common European Framework of Reference (CEFR).[25]

Second, vocabulary-based LexTALE[26] scores were collected from the Korean and Mandarin groups only. LexTALE scores were high (in the 60s) and did not differ significantly between groups (Welch-corrected two-sample $|t| = 0.410, p = 0.680$). The group means were consistent with “upper intermediate” (B2) proficiency in the CEFR. Thus, both proficiency metrics suggested that participants were relatively proficient users of English.

2.2 Materials

The speech materials for each language were based on dialogues in a play or drama: *Goncourt oder Die Abschaffung des Todes* for German,[27] *Coffee Prince* for Korean,[28] *Two Dogs' Opinions on Life* for Mandarin,[29] and *The Good Doctor* (“The Governess”, scene 3) for English.[30] The original text of each dialogue was edited to ensure that it was gender-neutral, emotionally neutral (e.g., by removing jokes), contemporary (e.g., by removing archaic words), without overly long turns, and representative of the phonemic inventory of the language.[31]

2.3 Procedure

The speaking task was completed in a sound-insulated room in London. Participants were instructed to read the two target dialogues naturally (i.e., not to put on an acting voice) and were seated in front of a microphone while facing the experimenter; the two went through each target dialogue together, with the participant reading one character's lines and the experimenter reading the other's lines. Recordings were made at 44.1 kHz with 16-bit resolution in stereo and were then converted to mono using Audacity.

Participants read the target dialogues in two drinking conditions (SOBER and INTOXICATED) in separate sessions on different days, no more than 14 days apart. They were instructed not to eat, drink, or use mouthwash in the two hours before each session and not to smoke in the half hour before each session. With the exception of the Korean speakers (who completed the conditions in the same order: SOBER, then INTOXICATED), the order in which the drinking conditions were completed was counterbalanced across participants. The LexTALE proficiency test was completed at the end of the SOBER condition.

In both conditions, participants' BAC was tested and monitored using a breathalyzer (AlcoMate Premium AL-7000). BAC was measured at the start of the session to ensure that participants came in with no alcohol in their system. In the INTOXICATED condition, participants consumed a predetermined amount of alcohol (vodka or rum, mixed with orange, lemon, or apple juice), estimated on the basis of their self-reported weight and BAC charts,[32] to reach a target BAC of 0.12%. Three-quarters of the alcohol was first poured into a glass; participants then decided on the amount of mixer and drank the mixture at their own pace. BAC was tested 15 minutes after the mixture was consumed and then every 3–5 minutes until it went over 0.12% and dropped back down to 0.12%. If BAC never got up to 0.12% at this point, a small top-up was given from the remaining alcohol. Once BAC had hit 0.12%, participants were taken into the recording room to complete the speaking task.

2.4 Analysis

For the purposes of analysis, each audio recording was divided into a set of utterances. An utterance was defined as a breath group, a stretch of speech often flanked by silent pauses and/or audible inhalations and often (but not always) coinciding with a sentence or clause. Given that speakers may exhibit a higher rate of disfluencies and speech errors when intoxicated,[33,34] the utterances were aurally inspected for disfluencies, speaker-generated noise, background noise, errors, and inaudibility. Utterances that contained one or more of the above issues were excluded from further analysis (such exclusions comprised 8–13% of all utterances across the three participant groups). If an utterance was produced multiple times consecutively (restarts), the last production was kept if it was free of errors.

Following aural inspection, utterances were subjected to acoustic analyses of f_0 and duration in Praat.[35] The f_0 analysis used the Praat function "To Pitch (cc)..." (cross-correlation), with a pitch floor and ceiling of 50 Hz and 300 Hz, respectively, and a time step of 0.01 sec. From Praat's voice report for a given utterance, a standard deviation (SD) of f_0 was extracted, yielding the dependent variable of f_0 variability, as well as a total duration value for the utterance. The final dataset submitted to statistical analysis comprised 17,083 data points (utterances/items): 3,742, 4,551, and 8,790, respectively, in the German, Korean, and Mandarin groups.[36]

The f_0 variability data were analyzed in four linear mixed-effects models using `lmerTest`[37] in R,[38] with sum coding of all categorical fixed effects.[39] Model 1, built on the L1 data, tested H1 and contained fixed effects for Group, Condition, and their interaction. Models 2–4, one model per L1 group, tested H2; each contained fixed effects for Language, Condition, and their interaction. Up to two control variables were also added to these models: Duration (msec; log-transformed to the base of 10 then z-transformed), which was added to all models, and Gender, which was added to all models except for the Korean group model (since all Korean participants were female). Duration was included to account for the possible dependence between f_0 variability and utterance duration.[40] Where relevant, Gender was also included as it is known to influence f_0 variability.[41] All models contained the maximal random-effects structure by Participant and Item.

All models underwent the process of model criticism.[42] For each model, the residuals were extracted and data points that were more than 2.5 SD above or below the mean residual value were excluded. This process resulted in no more than 2.1% of the data points being excluded from any of the models. Fixed-effect summaries of the final models can be found in the appendix (section 5), which shows model formulas in the table captions. Post hoc comparisons were carried out using `emmeans` (without p-value adjustment).[43]

3. Results

3.1 Question 1: Intoxication effects by L1 background

Median f_0 variability was higher in the intoxicated than the sober condition for all groups (Fig. 1). The intoxication effect differed across items, but a majority (62% for German, 57% for Korean, 56% for Mandarin) showed higher variability in the intoxicated condition.

Results of Model 1 partially supported H1: the effect of intoxication was indeed smaller (in fact, not significant) in Mandarin, but this was also the case in Korean. Model 1 indicated a significant Condition effect overall, with intoxicated speech showing higher-than-average variability ($\beta = 2.064, t = 3.514, p < 0.001$).[44] However, because interaction coefficients were negative, suggesting a reduced effect in Korean and Mandarin, we further

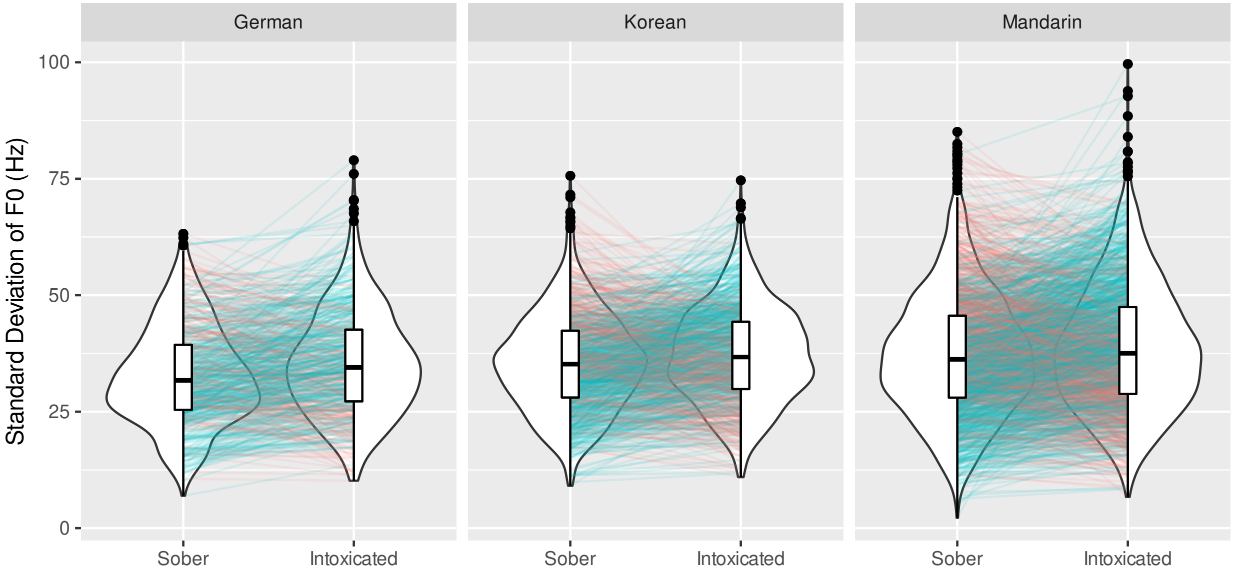

Fig. 1. Variability (SD) of f_0 in Hz in L1 utterances (items), by L1 group, condition, and item (horizontal lines). Blue indicates higher variability for the given item in the intoxicated condition.

inspected the magnitude of the intoxication effect (i.e., intoxicated - sober) by group/L1, finding a significant effect in German (estimate $= 3.232, z = 2.847, p = 0.004$) but not in Korean (estimate $= 1.690, z = 1.555, p = 0.120$) or Mandarin (estimate $= 1.269, z = 1.671, p = 0.095$). As always, null results should be interpreted cautiously; crucially, however, the null result (i.e., no intoxication effect) for Mandarin is consistent with H1. As for control predictors, there was a positive Duration effect ($\beta = 0.907, t = 2.796, p = 0.007$) and also a Gender effect whereby males showed lower-than-average variability ($\beta = -7.404, t = -3.909, p < 0.001$).

3.2 Question 2: Intoxication effects within the linguistic repertoire

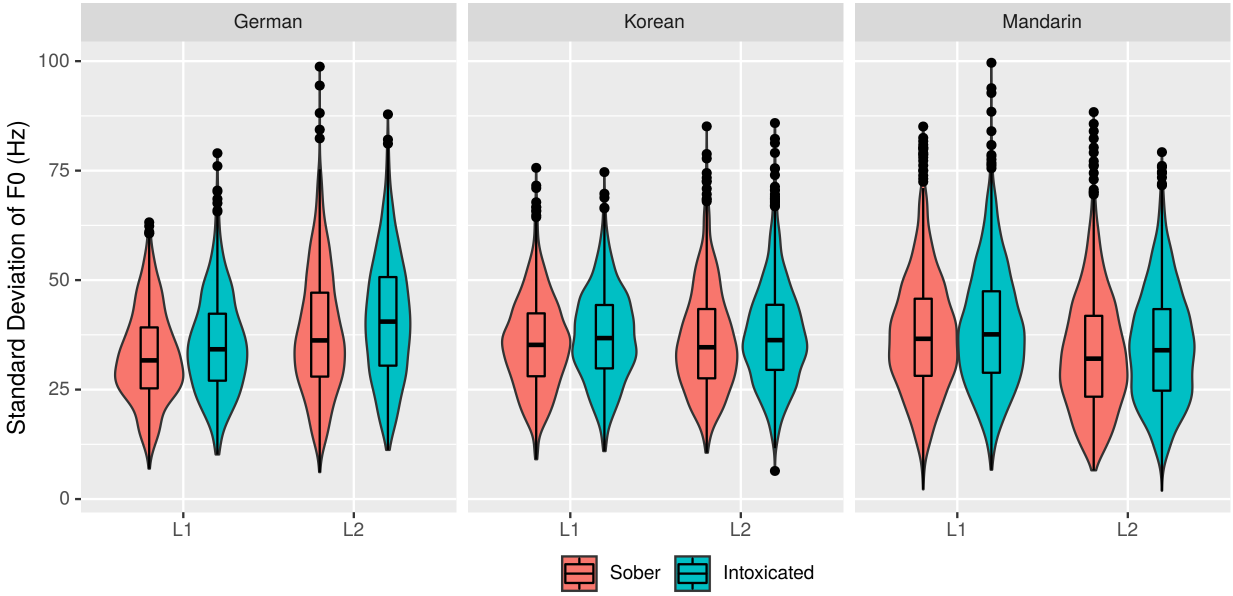

Fig. 2. Variability (SD) of f_0 in Hz, by L1 group, language (L1 or L2), and condition.

Median f_0 variability was higher in the intoxicated than the sober condition across all groups and languages (Fig. 2), but intoxication effects were largest in the German group. Results of Models 2–4 fully supported H2: for all groups, intoxication effects were similar between the L1 and L2. Inspection of intoxication effects by group and language revealed the same pattern in a group's L2 English as was observed in their L1: the German group showed a significant effect (estimate $= 2.960, z = 2.670, p = 0.008$), while the Korean (estimate $= 1.327, z = 0.716, p = 0.474$) and Mandarin (estimate $= 1.900, z = 1.708, p = 0.088$) groups did not. As for control predictors, there was no significant Duration effect in any model ($|\beta|$s $< 0.6, |t|$s $< 1.6, ps > 0.05$) and a significant Gender effect only in Model 4, whereby males showed lower-than-average variability as above ($\beta = -9.084, t = -4.078, p < 0.001$).

4. Discussion

This study directly compared the effects of intoxication on pitch control in speakers of tonal and non-tonal languages. It found evidence for a shared control mechanism for f_0 employed by bilinguals in their two languages: by allowing no significant increase in f_0 variability under intoxication, L1 speakers of Mandarin, a tonal language, showed greater overall control of f_0 variability in both the L1 and L2 (English), despite the fact that English is not a tonal language. Unexpectedly, greater overall pitch control was also found for L1 speakers of Korean, a non-tonal language; this may be related to a “quasi-tonal” prosodic system, in which there are no lexically specified tones but f_0 plays an important role in a limited set of phrasal positions as a cue to different consonantal laryngeal categories, which may in turn distinguish different lexical items. On the other hand, L1 speakers of German, a non-tonal language whose f_0 use is similar to that of English, showed less overall pitch control under intoxication in both the L1 and L2 data.

These findings have implications for phonetic typology as well as theories of bilingual phonology. First, while the results are compatible with the assertion that (Seoul) Korean is a “quasi-tonal” language, different types of languages verge on tonal (e.g., pitch-accent languages), and specific dialects may fall along a continuum of f_0 use, as has been shown for other languages (e.g., Basque, Japanese, Swedish).[45,46] In the case of Korean, there has been discussion about the status of some dialects as pitch-accent varieties, which points to the potential utility of intoxicated speech as a source of data on pitch control in speakers of understudied varieties. As above, null effects in this paradigm need to be interpreted cautiously, as they may arise for a number of reasons (e.g., individual differences in the effect of intoxication, socio-cultural factors related to appearing intoxicated); nevertheless, where intoxication consistently fails to affect speech production may turn out to be just as informative as where it does. Second, the current results support the view that bilingual phonological representations for pitch tend to be shared to some degree,[15–20] but more research is needed to understand the generalizability of these results to (psycho-)typologically different L1-L2 pairings. For instance, the consistent use of a non-tonal language as the L2 in the present study invites the question of what would happen when a tonal language is the L2. For example, might L1 English-L2 Mandarin speakers show, unlike L1 Mandarin-L2 English speakers, *less* overall pitch control under intoxication?

In closing, we would like to acknowledge two limitations of the current study, which point toward directions for future research. First, our findings are limited to read speech, which is known to show smaller effects of intoxication on f_0 properties than other speaking styles.[7] Therefore, it would be worthwhile to extend this work to diverse L1 populations producing a variety of speaking styles, including spontaneous speech. Second, this study leaves us with an incomplete picture of the role of gender, as our dataset did not allow an examination of gender effects in all groups. Given previous evidence of gender differences in f_0 modulation across languages,[17] it would thus be useful to further examine the effects of gender on f_0 variability. In addition, future research could explore correlations of f_0 variability changes with individual-difference variables (e.g., working memory), examine the effect of specific intonational tunes in our target dialogues on f_0 variability, and compare the effects of intoxication with other conditions known to affect speech, such as sleep deprivation.

5. Appendix

Table 1. Fixed effects in Model 1 (L1 data only). Model formula: `F0Var ~ Duration + Gender + Group + Condition + Group:Condition + (1 + Duration + Gender + Condition | Item) + (1 + Duration + Condition | Participant)`.

	β	SE	t	p	
(Intercept)	33.751	1.054	32.026	<0.001	***
Duration	0.907	0.324	2.796	0.007	**
Gender: male (vs. grand mean)	-7.404	1.894	-3.909	<0.001	***
Group: Korean (vs. grand mean)	-3.133	2.759	-1.136	0.264	
Group: Mandarin (vs. grand mean)	5.255	2.368	2.219	0.033	*
Condition: intoxicated (vs. grand mean)	2.064	0.587	3.514	<0.001	***
Group: Korean × Condition: intoxicated	-0.748	1.703	-0.439	0.664	
Group: Mandarin × Condition: intoxicated	-1.589	1.446	-1.099	0.280	

Observations: 7,822; participants: 33; items: 394.

Significance codes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2. Fixed effects in Model 2 (German group). Model formula: `F0Var` ~ `Duration + Gender + Language + Condition + Language:Condition + (1 + Duration + Gender + Condition | Item) + (1 + Duration + Language + Condition + Language:Condition | Participant)`.

	β	SE	t	p	
(Intercept)	35.174	2.740	12.836	<0.001	***
Duration	-0.577	0.550	-1.049	0.321	
Gender: male (vs. grand mean)	-4.611	2.581	-1.786	0.115	
Language: L2 (vs. grand mean)	5.465	1.664	3.284	0.012	*
Condition: intoxicated (vs. grand mean)	2.885	0.616	4.684	0.002	**
Language: L2 × Condition: intoxicated	0.154	1.288	0.120	0.908	

Observations: 3,661; participants: 8; items: 255.

Significance codes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3. Fixed effects in Model 3 (Korean group). Model formula: `F0Var ~ Duration + Language + Condition + Language:Condition + (1 + Duration + Condition | Item) + (1 + Duration + Language + Condition + Language:Condition | Participant)`.

	β	SE	t	p	
(Intercept)	36.227	1.045	34.666	<0.001	***
Duration	0.582	0.408	1.425	0.176	
Language: L2 (vs. grand mean)	0.193	1.577	0.122	0.906	
Condition: intoxicated (vs. grand mean)	1.561	1.105	1.413	0.200	
Language: L2 × Condition: intoxicated	-0.467	1.796	-0.261	0.802	

Observations: 4,476; participants: 8; items: 318.

Significance code: *** $p < 0.001$.

Table 4. Fixed effects in Model 4 (Mandarin group). Model formula: `F0Var` $\sim$ `Duration + Gender + Language + Condition + Language:Condition + (1 + Duration + Gender + Condition | Item) + (1 + Duration + Language + Condition + Language:Condition | Participant)`.

	β	SE	t	p	
(Intercept)	34.757	1.203	28.903	<0.001	***
Duration	0.516	0.324	1.594	0.121	
Gender: male (vs. grand mean)	-9.084	2.228	-4.078	<0.001	***
Language: L2 (vs. grand mean)	-5.121	1.031	-4.967	<0.001	***
Condition: intoxicated (vs. grand mean)	1.649	0.981	1.681	0.112	
Language: L2 × Condition: intoxicated	0.509	0.758	0.671	0.512	

Observations: 8,612; participants: 17; items: 295.

Significance code: *** $p < 0.001$.

Acknowledgments

The authors gratefully acknowledge data contributions from Ji Hye Kwon and Wen Jia Liu and helpful feedback from the Associate Editor and an anonymous reviewer.

References and links

[1] F. Trojan and K. Kryspin-Exner, “The decay of articulation under the influence of alcohol and paraldehyde,” Folia Phoniatrica et Logopaedica **20**(4), 217–238 (1968).

[2] S. L. Beam, R. W. Gant, and M. J. Mecham, “Communication deviations in alcoholics; A pilot study,” Journal of Studies on Alcohol and Drugs **39**(3), 548–551 (1978).

[3] S. B. Chin and D. B. Pisoni, *Alcohol and Speech* (Academic Press, San Diego, CA, 1997).

[4] D. B. Pisoni and C. S. Martin, “Effects of alcohol on the acoustic-phonetic properties of speech: Perceptual and acoustic analyses,” Alcoholism: Clinical and Experimental Research **13**(4), 577–587 (1989).

[5] K. Johnson, D. B. Pisoni, and R. H. Bernacki, “Do voice recordings reveal whether a person is intoxicated? A case study,” Phonetica **47**(3–4), 215–237 (1990).

[6] H. Hollien, G. DeJong, C. A. Martin, R. Schwartz, and K. Liljegren, “Effects of ethanol intoxication on speech suprasegmentals,” Journal of the Acoustical Society of America **110**(6), 3198–3206 (2001).

[7] B. Baumeister, C. Heinrich, and F. Schiel, “The influence of alcoholic intoxication on the fundamental frequency of female and male speakers,” Journal of the Acoustical Society of America **132**(1), 442–451 (2012).

[8] H. Watanabe, T. Shin, H. Matsuo, F. Okuno, T. Tsuji, M. Matsuoka, J. Fukaura, and H. Matsunaga, “Studies on vocal fold injection and changes in pitch associated with alcohol intake,” Journal of Voice **8**(4), 340–346 (1994).

[9]A. Z. Guiora, B. Beit-Hallahmi, R. C. L. Brannon, C. Y. Dull, and T. Scovel, “The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study,” Comprehensive Psychiatry **13**(5), 421–428 (1972).

[10]M. Wieling, C. Blankevoort, V. Hukker, J. Jacobi, L. de Jong, S. Keulen, M. Medvedeva, M. van der Ploeg, A. Pot, T. Rebernik, P. Veenstra, and A. Noiray, “The influence of alcohol on L1 vs. L2 pronunciation,” in *Proceedings of the 19th International Congress of Phonetic Sciences*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc., Canberra, Australia, 2019), pp. 3622–3626.

[11]C. B. Chang, “Determining cross-linguistic phonological similarity between segments: The primacy of abstract aspects of similarity,” in *The Segment in Phonetics and Phonology*, edited by E. Raimy and C. E. Cairns (John Wiley & Sons, Chichester, UK, 2015), pp. 199–217.

[12]C. B. Chang, “The phonetics of second language learning and bilingualism,” in *The Routledge Handbook of Phonetics*, edited by W. F. Katz and P. F. Assmann (Routledge, Abingdon, UK, 2019), pp. 427–447.

[13]C. B. Chang, “Phonetic drift,” in *The Oxford Handbook of Language Attrition*, edited by M. S. Schmid and B. Köpke (Oxford University Press, Oxford, UK, 2019), pp. 191–203.

[14]E. de Leeuw, “Phonetic L1 attrition,” in *The Oxford Handbook of Language Attrition*, edited by M. S. Schmid and B. Köpke (Oxford University Press, Oxford, UK, 2019), pp.

204–217.

[15] I. Mennen, "Bi-directional interference in the intonation of Dutch speakers of Greek," Journal of Phonetics **32**(4), 543–563 (2004).

[16] E. de Leeuw, I. Mennen, and J. M. Scobbie, "Singing a different tune in your native language: First language attrition of prosody," International Journal of Bilingualism **16**(1), 101–116 (2012).

[17] M. Ordin and I. Mennen, "Cross-linguistic differences in bilinguals' fundamental frequency ranges," Journal of Speech, Language, and Hearing Research **60**(6), 1493–1506 (2017).

[18] C. B. Chang, "Rapid and multifaceted effects of second-language learning on first-language speech production," Journal of Phonetics **40**(2), 249–268 (2012).

[19] C. B. Chang, "A novelty effect in phonetic drift of the native language," Journal of Phonetics **41**(6), 491–504 (2013).

[20] C. B. Chang, "Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research," Journal of Phonetics **74**, 96–113 (2019).

[21] Y. Kang, "Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study," Journal of Phonetics **45**, 76–90 (2014).

[22] H.-Y. Bang, M. Sonderegger, Y. Kang, M. Clayards, and T.-J. Yoon, "The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis," Journal of Phonetics **66**, 120–144 (2018).

[23] C. B. Chang, “Perceptual attention as the locus of transfer to nonnative speech perception,” Journal of Phonetics **68**, 85–102 (2018).

[24] D. Surendran and G.-A. Levow, “The functional load of tone in Mandarin is as high as that of vowels,” in *Proceedings of the 2nd International Conference on Speech Prosody*, edited by B. Bel and I. Marlien (International Speech Communication Association, Nara, Japan, 2004), pp. 99–102.

[25] See the IELTS website: https://www.ielts.org/about-ielts/ielts-in-cefr-scale.

[26] K. Lemhöfer and M. Broersma, “Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English,” Behavior Research Methods **44**(2), 325–343 (2012).

[27] T. Dorst and H. Laube, *Goncourt oder Die Abschaffung des Todes* (Schauspiel Frankfurt, Frankfurt, Germany, 1977).

[28] J.-A. Lee and H.-J. Jang, “Keopi peurinseu 1 hojeom [The 1st shop of coffee prince],” (2007), South Korean TV series, directed by Yoon-Jung Lee and premiered on Munhwa Broadcasting Corporation (MBC).

[29] J. Meng, “Liǎng zhī gǒu de shēnghuó yìjiàn [Two dogs’ opinions on life],” (2008), Chinese theatrical play, presented by the National Theatre of China.

[30] N. Simon, *The Collected Plays of Neil Simon: Volume 2* (Plume, New York, NY, 1986).

[31] All materials are available open-access at https://osf.io/y2r87/.

[32] W. R. Miller and R. F. Muñoz, *How to Control Your Drinking* (Prentice Hall, Englewood Cliffs, NJ, 1982).

[33] F. Schiel and C. Heinrich, “Disfluencies in the speech of intoxicated speakers,” International Journal of Speech Language and the Law **22**(1), 19–34 (2015) doi: `10.1558/ijsll.v22i1.24767`.

[34] A. Cutler and C. G. Henton, “There’s many a slip ’twixt the cup and the lip,” in *On Speech and Language: Studies for Sieb G. Nooteboom*, edited by H. Quené and V. van Heuven (Landelijk Onderzoekschool Taalwetenschap, Utrecht, 2004), pp. 37–45.

[35] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer,” (2016), version 6.0.19. http://www.praat.org.

[36] The dataset is available open-access at `https://osf.io/d9e78/`.

[37] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmertest package: Tests in linear mixed effects models,” Journal of Statistical Software **82**(13), 1–26 (2017).

[38] R Development Core Team, “R: A language and environment for statistical computing,” (2018), version 3.5.2. http://www.r-project.org.

[39] M. Wissmann, H. Toutenburg, and Shalabh, “Role of categorical variables in multicollinearity in the linear regression,” Technical Report 008, Department of Statistics, University of Munich, Munich, Germany (2007), `https://epub.ub.uni-muenchen.de/2081/1/report008_statistics.pdf`.

[40] This is because the longer the utterance duration, the more f_0 observations there will be within that utterance. On the one hand, more observations could tend to lower variability, because the number of observations sits in the denominator of the SD formula. On the other

hand, more observations could also raise variability, because they introduce more chances for divergence from the grand mean, which sits in the numerator of the SD formula.

[41] A. P. Simpson, “Phonetic differences between male and female speech,” Language and Linguistics Compass **3**(2), 621–640 (2009).

[42] R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge, UK, 2008).

[43] R. V. Lenth, P. Buerkner, M. Herve, J. Love, H. Riebl, and H. Singmann, “emmeans: Estimated marginal means, aka least-squares means [R package],” (2021), version 1.7.0. Available from https://cran.r-project.org/web/packages/emmeans/index.html.

[44] As pointed out by an anonymous reviewer, Model 1 also suggested that f_0 variability in Mandarin (averaging over both conditions) was significantly higher than the grand mean over all L1s. Although this result is not directly relevant to H1, it may be related to the relatively high number of different pitch shapes in Mandarin (four main contrastive tones), which could have the effect of increasing f_0 variability within any given utterance.

[45] M. Shibatani, *The Languages of Japan* (Cambridge University Press, Cambridge, UK, 1990).

[46] J. I. Hualde, “Two Basque accentual systems and the notion of pitch-accent language,” Lingua **122**(13), 1335–1351 (2012).