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 Nevins3

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
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.

© 2022 Acoustical Society of America.

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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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. 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.

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. 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, $f_0$ range for L1 Welsh-L2 English speakers (albeit mostly for males), and
f₀ level for L1 English-L2 Korean speakers, consistent with the view that there is a crosslinguistically “shared control mechanism for f₀ modulation.” Few studies, however, have examined f₀ 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), 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), 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. This leads to the hypothesis that intoxication will impact the variability of f₀ (the acoustic correlate of pitch) less for L1 Mandarin speakers than for L1 German or Korean speakers (H1). Furthermore, if f₀ 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 ($N = 8; 4f, 4m; M_{age} = 27.1 \text{ yr}, SD = 4.3$), Korean ($N = 8; 8f, 0m; M_{age} = 27.1 \text{ yr}, SD = 3.8$), and Mandarin ($N = 17; 10f, 7m; M_{age} = 23.8 \text{ yr}, SD = 1.5$) 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 $|t|s < 1.7, ps > 0.05$). 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).  

Second, vocabulary-based LexTALE 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, *Coffee Prince* for Korean, *Two Dogs’ Opinions on Life* for Mandarin, and *The Good Doctor* (“The Governess”, scene 3) for English. 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.

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, 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,\textsuperscript{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.\textsuperscript{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.\textsuperscript{36}

The $f_0$ variability data were analyzed in four linear mixed-effects models using lmerTest\textsuperscript{37} in R,\textsuperscript{38} with sum coding of all categorical fixed effects.\textsuperscript{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. Where relevant, Gender was also included as it is known to influence $f_0$ variability. All models contained the maximal random-effects structure by Participant and Item.

All models underwent the process of model criticism. 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).

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 ($\beta = 2.064, t = 3.514, p < 0.001$). 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
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). 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, 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. 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, 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: $\text{F0Var} \sim \text{Duration} + \text{Gender} + \text{Group} + \text{Condition} + \text{Group:Condition} + (1 + \text{Duration} + \text{Gender} + \text{Condition} | \text{Item}) + (1 + \text{Duration} + \text{Condition} | \text{Participant})$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$SE$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>33.751</td>
<td>1.054</td>
<td>32.026</td>
<td>$&lt;0.001^{***}$</td>
</tr>
<tr>
<td>Duration</td>
<td>0.907</td>
<td>0.324</td>
<td>2.796</td>
<td>0.007 $^{**}$</td>
</tr>
<tr>
<td>Gender: male (vs. grand mean)</td>
<td>-7.404</td>
<td>1.894</td>
<td>-3.909</td>
<td>$&lt;0.001^{***}$</td>
</tr>
<tr>
<td>Group: Korean (vs. grand mean)</td>
<td>-3.133</td>
<td>2.759</td>
<td>-1.136</td>
<td>0.264</td>
</tr>
<tr>
<td>Group: Mandarin (vs. grand mean)</td>
<td>5.255</td>
<td>2.368</td>
<td>2.219</td>
<td>0.033 $^{*}$</td>
</tr>
<tr>
<td>Condition: intoxicated (vs. grand mean)</td>
<td>2.064</td>
<td>0.587</td>
<td>3.514</td>
<td>$&lt;0.001^{***}$</td>
</tr>
<tr>
<td>Group: Korean $\times$ Condition: intoxicated</td>
<td>-0.748</td>
<td>1.703</td>
<td>-0.439</td>
<td>0.664</td>
</tr>
<tr>
<td>Group: Mandarin $\times$ Condition: intoxicated</td>
<td>-1.589</td>
<td>1.446</td>
<td>-1.099</td>
<td>0.280</td>
</tr>
</tbody>
</table>

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

Significance codes: $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. 15
Table 2. Fixed effects in Model 2 (German group). Model formula: $F_0\text{Var} \sim \text{Duration} + \text{Gender} + \text{Language} + \text{Condition} + \text{Language:Condition} + (1 + \text{Duration} + \text{Gender} + \text{Condition} | \text{Item}) + (1 + \text{Duration} + \text{Language} + \text{Condition} + \text{Language:Condition} | \text{Participant})$.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>35.174</td>
<td>2.740</td>
<td>12.836</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>Duration</td>
<td>-0.577</td>
<td>0.550</td>
<td>-1.049</td>
<td>0.321</td>
</tr>
<tr>
<td>Gender: male (vs. grand mean)</td>
<td>-4.611</td>
<td>2.581</td>
<td>-1.786</td>
<td>0.115</td>
</tr>
<tr>
<td>Language: L2 (vs. grand mean)</td>
<td>5.465</td>
<td>1.664</td>
<td>3.284</td>
<td>0.012*</td>
</tr>
<tr>
<td>Condition: intoxicated (vs. grand mean)</td>
<td>2.885</td>
<td>0.616</td>
<td>4.684</td>
<td>0.002**</td>
</tr>
<tr>
<td>Language: L2 $\times$ Condition: intoxicated</td>
<td>0.154</td>
<td>1.288</td>
<td>0.120</td>
<td>0.908</td>
</tr>
</tbody>
</table>

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: \( \text{F0Var} \sim \text{Duration} + \text{Language} + \text{Condition} + \text{Language}:\text{Condition} + (1 + \text{Duration} + \text{Condition} | \text{Item}) + (1 + \text{Duration} + \text{Language} + \text{Condition} + \text{Language}:\text{Condition} | \text{Participant}) \).

<table>
<thead>
<tr>
<th></th>
<th>( \beta )</th>
<th>( SE )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>36.227</td>
<td>1.045</td>
<td>34.666</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Duration</td>
<td>0.582</td>
<td>0.408</td>
<td>1.425</td>
<td>0.176</td>
</tr>
<tr>
<td>Language: L2 (vs. grand mean)</td>
<td>0.193</td>
<td>1.577</td>
<td>0.122</td>
<td>0.906</td>
</tr>
<tr>
<td>Condition: intoxicated (vs. grand mean)</td>
<td>1.561</td>
<td>1.105</td>
<td>1.413</td>
<td>0.200</td>
</tr>
<tr>
<td>Language: L2 \times Condition: intoxicated</td>
<td>-0.467</td>
<td>1.796</td>
<td>-0.261</td>
<td>0.802</td>
</tr>
</tbody>
</table>

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

Significance code: *** \( p < 0.001 \).
Table 4. Fixed effects in Model 4 (Mandarin group). Model formula: \( \text{F0Var} \sim \text{Duration} + \text{Gender} + \text{Language} + \text{Condition} + \text{Language:Condition} + (1 + \text{Duration} + \text{Gender} + \text{Condition} | \text{Item}) + (1 + \text{Duration} + \text{Language} + \text{Condition} + \text{Language:Condition} | \text{Participant}) \).

<table>
<thead>
<tr>
<th></th>
<th>( \beta )</th>
<th>SE</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>34.757</td>
<td>1.203</td>
<td>28.903</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Duration</td>
<td>0.516</td>
<td>0.324</td>
<td>1.594</td>
<td>0.121</td>
</tr>
<tr>
<td>Gender: male (vs. grand mean)</td>
<td>-9.084</td>
<td>2.228</td>
<td>-4.078</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Language: L2 (vs. grand mean)</td>
<td>-5.121</td>
<td>1.031</td>
<td>-4.967</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Condition: intoxicated (vs. grand mean)</td>
<td>1.649</td>
<td>0.981</td>
<td>1.681</td>
<td>0.112</td>
</tr>
<tr>
<td>Language: L2 x Condition: intoxicated</td>
<td>0.509</td>
<td>0.758</td>
<td>0.671</td>
<td>0.512</td>
</tr>
</tbody>
</table>

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


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


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


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


The dataset is available open-access at https://osf.io/d9e78/.


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

42 R. H. Baayen, Analyzing Linguistic Data: A Practical Introduction to Statistics Using R

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,”