Ganong effects for frequency may not be robust

Stephen Politzer-Ahles, Ka Keung Lee, and Lue Shen

ARTICLES YOU MAY BE INTERESTED IN

Numerical modeling of submicron particles for acoustic concentration in gaseous flow
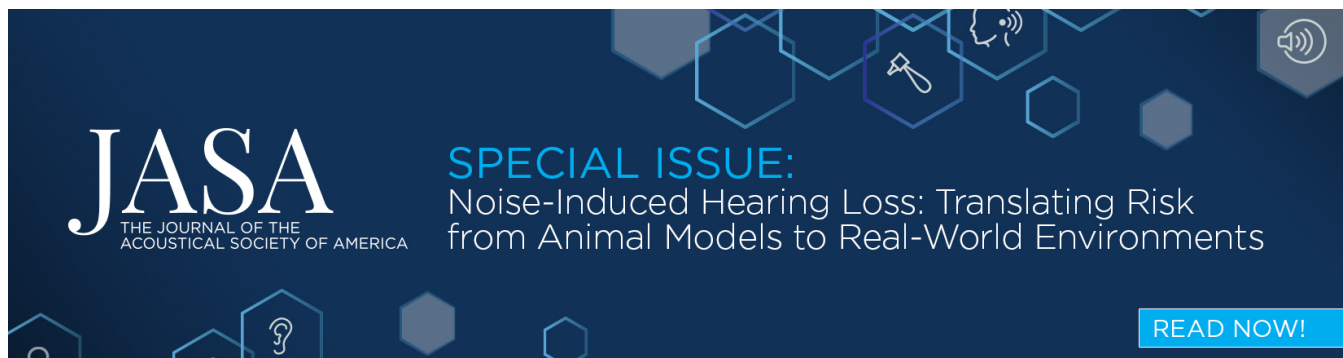The Journal of the Acoustical Society of America **147**, 152 (2020); https://doi.org/10.1121/10.0000568

Evidence supporting synchrony between two active ears due to interaural coupling
The Journal of the Acoustical Society of America **147**, EL25 (2020); https://doi.org/10.1121/10.0000473

Reflection of an ultrasonic wave on the bone-implant interface: Comparison of two-dimensional and three-dimensional numerical models
The Journal of the Acoustical Society of America **147**, EL32 (2020); https://doi.org/10.1121/10.0000500

Ganong effects for frequency may not be robust

Stephen Politzer-Ahles,[1,a] Ka Keung Lee,[1] and Lue Shen[2]

[1]*Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong*
[2]*Program in Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts 02215,*
USA
stephen.politzerahles@polyu.edu.hk, leon-leechk@hotmail.com, shenlue@bu.edu

Abstract: The Ganong effect—more identifications of a certain phoneme in a context where that phoneme would yield a real word than a context where that phoneme would yield a pseudoword—has been widely replicated. Few studies, however, have tested whether this effect occurs for frequency contrasts. In the present study, participants' likelihood of identifying an ambiguous sound as aspirated was tested in acoustically identical continua in contexts where the identification of the sound as aspirated would either yield a lower- or higher-frequency word than the identification of the sound as unaspirated would. No frequency-based Ganong effect was found.

1. Introduction

When hearing an ambiguous sound, listeners' identification of the sound may be influenced by whether or not it forms a real word in its context (Ganong, 1980). For example, when hearing a sound that is ambiguous between [d] and [t], English speakers tend to be more likely to identify the sound as [t] when in the context _ask (as *task* is a word and **dask* is not) compared to the context _esk (as **tesk* is not a word and *desk* is). This Ganong effect has been widely replicated, although the specific locus and mechanism of the effect remains under debate (see, e.g., Kingston *et al.*, 2016).

The Ganong effect has typically been tested using manipulations comparing contexts where one interpretation of the ambiguous sound yields a word and the other a pseudoword (e.g., *task* and **dask*, **tesk* and *desk*). Most accounts of why the Ganong effect occurs, however, would also predict a similar effect to arise if different interpretations of a sound yield higher- and lower-frequency words. For instance, a sound ambiguous between [d] and [t] might be more likely to be identified as [t] when it is heard in the context _ime (given that *time* is a more common word than *dime*) compared to the context _or (given that *tore* is a less common word than *door*). Thus far, only one published study, by Connine and colleagues (1993), has tested this situation. They found the expected frequency-based Ganong effect.

Given the importance of the Ganong effect for models of speech perception (see, e.g., Norris *et al.*, 2000, among others) and the dearth of experimental evidence for frequency-based Ganong effects, we felt it would be valuable to attempt to replicate and extend these findings to another language. In addition to the inherent value of replicating important results, testing this pattern in Chinese is of particular value (as noted by an anonymous reviewer) since the psychological importance of segments may be different in Chinese than it is in Indo-European languages (see, e.g., O'Seaghdha *et al.*, 2010). Furthermore, in the present study we adopt a different, and possibly more controlled, method for creating experimental stimuli. Connine and colleagues (1993) made continua of ambiguous stimuli by successively cutting out or including longer portions of periodic or aperiodic energy from natural stimuli (e.g., to create a continuum from *dime* to *time*, successively longer portions of the prevoicing from *dime* were removed, while at the same time successively longer portions of aspiration from *time* were spliced in). This method may introduce differences between items/continua, since aspiration is not uniform and thus cutting out 10 ms (for example) from one portion of the aspirated period may have different effects than cutting out 10 ms from another portion, and these differences may not be matched across different continua (since the *dime–time* continuum is edited separately from, e.g., the *door–tore* continuum). Furthermore, the onset and offset of aspiration (or prevoicing) are not always obvious, and thus stimuli treated by the experimenter as having, for example, 20 ms of voice onset time

[a]Author to whom correspondence should be addressed, ORCID: 0000-0002-5474-7930.

(VOT) may actually have 19 ms in one continuum, 21 ms in another continuum, etc., depending on small variations where the cursor may be placed when measuring VOT. The stimuli used by Connine and colleagues (1993) were manipulated very carefully and systematically, and the number of items was large enough that these kinds of small and probably random differences are not likely to have introduced systematic confounds in the results. Nevertheless, we felt it would be valuable to attempt to find converging evidence for frequency-based Ganong effects using stimuli in which the ambiguous portions are completely acoustically identical across different continua.

To accomplish this, we used two-syllable Mandarin words, where the first syllable is always the same (other than VOT of its onset consonant) across continua, and the lexicality or frequency of the continuum endpoints is modulated by the second syllable. For example, *duìhuà* ("conversation") is a higher-frequency word than *tuìhuà* ("degeneration"), whereas *duìyì* ("play chess") is a lower-frequency word than *tuìyì* ("retire"). Crucially, all of these words start with *duì/tuì*, and thus *duìhuà–tuìhuà* and *duìyì–tuìyì* continua can be created by using the same *duì–tuì* continuum and splicing different second syllables (*huà* or *yì*) onto the end of each step.

A preliminary study (Shen and Politzer-Ahles, 2018) in this design using only one set of items (the *duì–tuì* continuum described above, with additional control conditions where the [d] end of the continuum forms a real word and the [t] end a pseudoword, where the [d] end forms a pseudoword and [t] a word, and when both ends form a pseudoword) and 35 participants found the expected effects for both lexicality and frequency. However, a limitation of this design compared to Connine and colleagues' (1993) is that the results are based on just one set of continua, and thus may be influenced by other factors we have not considered (see, e.g., Norris *et al.*, 2000, Sec. 4.3, for a powerful demonstration of how effects that seem to be due to lexicality may turn out to be due to lower-level factors like phoneme transitional probabilities), whereas Connine and colleagues' (1993) design uses a large number of items and thus minimizes the potential for such confounds. Thus, in the present study, we used a controlled design like that described above, but included a larger set of items, in order to test whether a frequency-based Ganong effect can be found to generalize across items in this paradigm.

2. Methods

All data, stimuli, and analysis codes are available at https://osf.io/8y94s/ (experiment 2). The experiment methods were pre-registered at https://osf.io/6e35g.

2.1 Participants

Seventy native speakers of Mandarin (aged 19–34 years, mean 24 years, 53 women and 12 men, as well as five who chose not to indicate their gender) took part in the experiment in Hong Kong. Experimental procedures were approved by the Human Subjects Ethics Sub-Committee at the Hong Kong Polytechnic University. Participants provided informed consent and were reimbursed with cash for their participation.

2.2 Stimuli

We chose three sets of stimuli (i.e., three items) with onset stops from different places of articulation. For each item, the first syllable was always the same, and four different second syllables were used to yield four contexts, as shown in Table 1. The contexts in which treating the onset stop as aspirated yields a word or a pseudoword are the classic Ganong effect manipulation. The contexts in which treating the onset stop as aspirated yield a higher- or lower-frequency word [according to the SUBTLEX-CH corpus (Cai and Brysbaert, 2010)] are the manipulation for testing whether word frequency also yields a Ganong effect.

The stimuli were recorded by a female Mandarin native speaker (aged 19) from Harbin. Each first syllable (*pì*, *tàn*, and *kuān*) was manipulated using Praat (Boersma and Weenink, 2017) into an unaspirated-to-aspirated continuum (5 to 95 ms of voice onset time, in 5-ms steps) and a discrimination pretest with four participants was run to estimate the location of the categorical boundary in each continuum. In the real experiment, for each item we used a nine-step continuum centered on that categorical boundary (25 ms for *pì* and *tàn*, 50 ms for *kuān*) with 5-ms increments between steps. Finally, each first syllable was spliced to each of the four possible second syllables, yielding 108 tokens (9 steps * 3 items * 4 contexts).

2.3 Procedure

Stimulus presentation and response logging were controlled using DMDX (Forster and Forster, 2003; stimuli and scripts available at https://osf.io/8y94s/). The experiment included nine blocks, and each block included stimuli from just one item (*pì*, *tàn*, or *kuān* stimuli). In a given block, all 36 tokens for that stimulus (9 steps * 4 contexts) were repeated once in a random order, and then one more time in a random order. Each item occurred in three blocks; therefore, each token

Table 1. Stimuli used in the experiment. The first row of each cell gives the Hanyu Pinyin transcription of the two ends of the continuum, the second row gives the frequencies [log words per million, from the SUBTLEX-CH corpus (Cai and Brysbaert, 2010)] of the words at the continuum endpoints, and the third row gives the corresponding log words-per-million frequencies from the Sinica Corpus 4.0 (Chen *et al.*, 1996; http://asbc.iis.sinica.edu.tw/). Where the word does not appear in the corpus, the log frequency is NA. While *dànwàng* does not appear in the SUBTLEX-CH corpus, it was identified as a recognizable word by the authors and our informants (see, e.g., https://baike.baidu.com/item/%E6%B7%A1%E5%BF%98/2915944). For the first row of the velar continua, we report the frequency for 宽敞, which SUBTLEX-CH lists as *kuānchang* (i.e., with a neutral tone rather than a Tone 3 on the final syllable), but which is also commonly pronounced *kuānchǎng*. The Academica Sinica corpus has fewer tokens of *pìhuà* than *bìhuà*, contrary to our assumption that *pìhuà* is higher frequency; we suspect that this is because that corpus consists mostly of written and relatively formal texts (from categories including literature, lifestyle, society, science, philosophy, and art) whereas *pìhuà* ("bullshit") mainly occurs in speech or less formal writing (e.g., web forums). The SUBTLEX-CH corpus consists of movie subtitles, and thus has more tokens of informal items like this.

	bilabial	alveolar	velar
Aspirated yields a higher-frequency word than unaspirated does	{p/b}ìhuà 1.8, 0.7 −1.7, 2.3	{t/d}ànwàng 1.9, NA 2.3, 0.8	{k/g}uānchǎng 1.1, -2.3 2.0, 1.6
Aspirated yields a lower-frequency word than unaspirated does	{p/b}ìjǐng −1.2, 3.1 0.5, 4.1	{t/d}ànshì 1.4, 7.4 2.1, 6.6	{k/g}uānxīn 0, 4.8 −0.2, 4.7
Aspirated yields a real word and unaspirated a pseudoword	{p/b}ìrú 0.7, NA 4.1, NA	{t/d}ànsuǒ 3.1, NA 3.3, NA	{k/g}uānróng 1.8, NA 2.0, NA
Aspirated yields a pseudoword and unaspirated a real word	{p/b}ìmiǎn NA, 3.3 NA, 5.0	{t/d}àngāo NA, 4.1 NA, 2.4	{k/g}uāndiǎn NA, 3.4 NA, 4.6

was responded to six times (2 repetitions * 3 blocks). The order of the blocks was random. Participants completed 648 trials (108 tokens * 6 repetitions). On each trial, they heard the token and then were asked to identify the first syllable. The prompt showed both options written in Hanyu Pinyin without tones (e.g., "<- bi … pi ->" or "<- guan … kuan ->"); the unaspirated option was always on the left. The prompt appeared on screen at the same time the stimulus played, and participants had 8 s from the stimulus onset to make their response before the program proceeded to the next trial. Participants were given self-paced breaks between blocks.

2.4 Analysis

Trials that timed out, or with reaction times below 200 ms (a common cutoff used in [typically visual] lexical decision studies, based on the assumption that responses this fast must be mistakes), were excluded from analysis. Statistical analysis was conducted using generalized (logistic) linear mixed-effects models (Baayen *et al.*, 2008) with random effects for participants; maximal random slopes justified by the design were used (Barr *et al.*, 2013). We also report percentile bootstrap confidence intervals (CIs) for the size of the Ganong effect in each comparison (expressed as the percentage of aspirated responses at the middle VOT step in the aspirated-biased continuum, minus that in the unaspirated-biased continuum). The models were fitted using the {lme4} package (Bates *et al.*, 2015) of the R statistical computing environment (R Core Team, 2016).

3. Results

All data files are available at https://osf.io/8y94s/. The average reaction times for stimuli in the lexicality manipulation were 982 ms for stimuli on the unaspirated-biased continuum and 1017 ms on the aspirated-biased continuum; for the frequency manipulation the average reaction times were 991 and 1000 ms, respectively, on the unaspirated- and aspirated- biased continua.

Figure 1 shows the results of the experiment; the top row shows the overall averages, and the second row shows the Ganong effect for each participant, along with the 95% percentile bootstrap CI of the effect. There was a strong Ganong effect for the comparison between the word-pseudoword continua, with more aspirated responses in the aspirated-biased continuum than in the unaspirated-biased continuum ($b = 0.70$, $z = 6.49$, $p < 0.001$, 95% CI = [18.6, 29.1]). On the other hand, there was no clear Ganong effect for the comparison between the higher- vs lower-frequency continua ($b < 0.01$, $z = 0.10$, $p = 0.917$, 95% CI = −3.7, 1.5]). The CI shows that not only was the Ganong effect for frequency continua not significantly different from zero, but it was also significantly *smaller* than the effect for lexicality continua. In other words, it is not
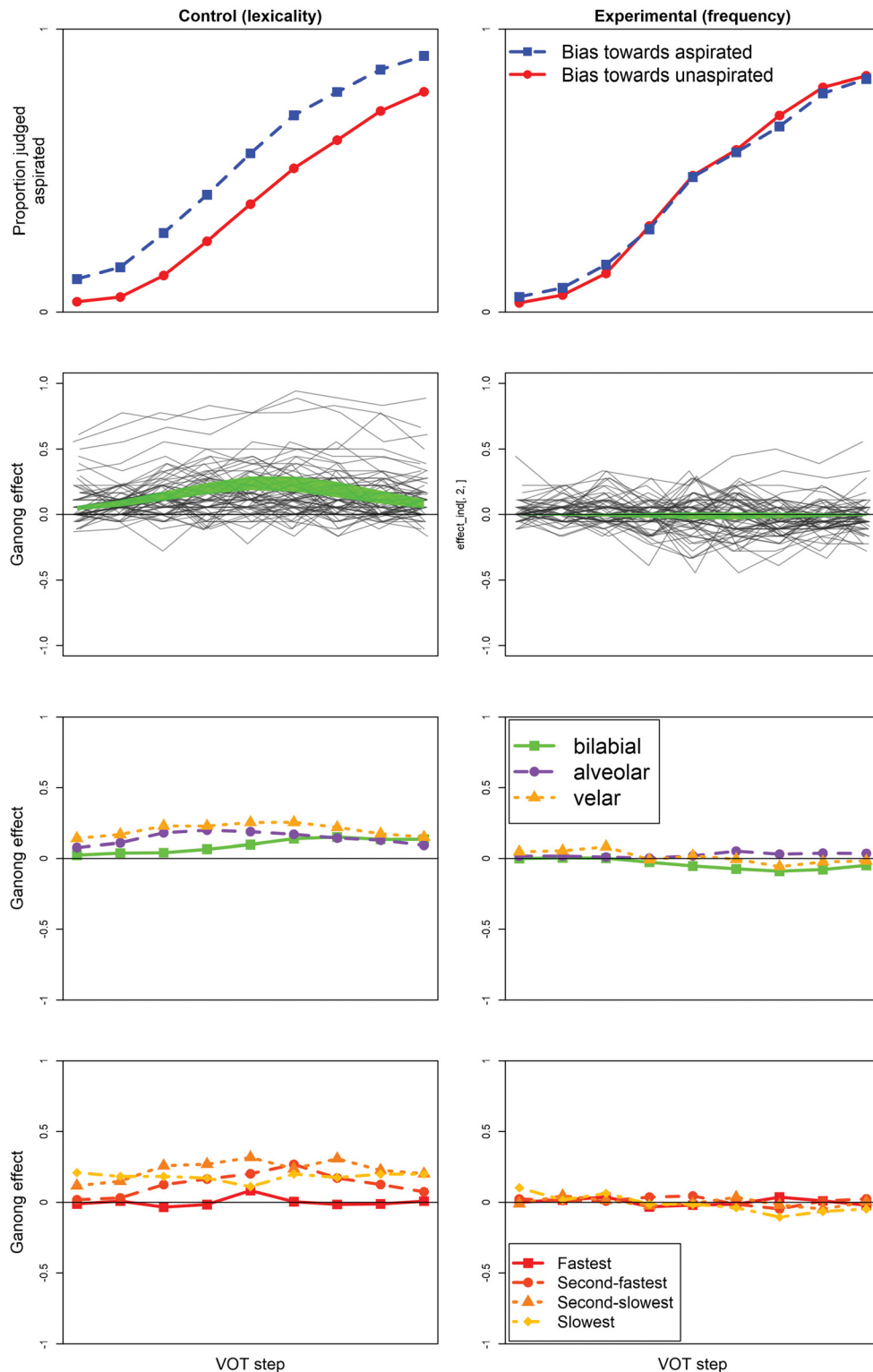
Fig. 1. (Color online) Results from the experiment. (A) The top row shows results averaged across participants, and the second row shows results for all individual participants. The left column shows results for the comparison between word-pseudoword continua, and the right column shows results for the comparison between higher- vs lower-frequency continua. In the aggregate results, the dashed line indicates the proportion of aspirated responses (per continuum step) in the context where lexicality or frequency should bias the participant to choose the aspirated response; the solid line indicates proportion of aspirated responses in the context where lexicality or frequency should bias the participant to choose the unaspirated response. (B) The second row shows the results for individual participants, where each thin line shows the Ganong effect (proportion of aspirated responses in the aspirated-biasing context, minus proportion of aspirated responses in the unaspirated-biasing context) for an individual participant, such that lines above zero are consistent with Ganong effects. The shaded areas are 95% CIs of the Ganong effect. (C) The third row shows the Ganong effect for each item (each continuum pair), following the same format as the individual-participants graph in the above row. (D) The fourth row shows the Ganong effect for each reaction time quartile.

the case that the experiment lacked statistical power to detect small Ganong effects for frequency; even small or negligible effects (on the order of two percentage points) are ruled out by these data. A model testing the interaction between contextual bias (aspirated- vs unaspirated-biased continua) and bias source (lexicality vs frequency) also found that the [lack of] Ganong effect for frequency was significantly smaller than the Ganong effect for lexicality ($b = -1.06$, $z = -8.30$, $p < 0.001$).

We also conducted exploratory analyses to test the possibility that maybe the expected Ganong effects do occur, but only in certain items or only at certain reaction times. The size of the frequency-based Ganong effect, however, was not reliably moderated by reaction time ($b = -0.03$, $z = -0.79$, $p = 0.430$); this observation is consistent with the findings of Connine and colleagues (1993). On the other hand, the size of the frequency-based Ganong effect was significantly moderated by stimulus [$\chi^2(2) = 11.11$, $p = 0.004$]; the effect was not significant in the alveolar ($b = 0.18$, $z = 1.52$, $p = 0.130$) or velar items ($b = 0.07$, $z = 0.52$, $p = 0.602$), but was significant in the wrong direction (more "aspirated" responses in the unaspirated-biased context) in the labial items ($b = -0.46$, $z = -3.26$, $p = 0.001$). In other words, it does not seem to be the case that we would have obtained a frequency Ganong effect if we only focused on the "best" item or excluded the "worst"; rather, we do not obtain the effect in any pair of continua. (Removing the bilabial pair of continua, which had an effect in the wrong direction, we still do not obtain a significant Ganong effect in a combined analysis of the other two pairs: $b = 0.09$, $z = 1.17$, $p = 0.244$.)

4. Discussion

Unlike previous studies (Connine *et al.*, 1993; Shen and Politzer-Ahles, 2018), we found no trace of a frequency-based Ganong effect. Our sample size (in terms of the number of participants and number of trials, although not number of items) was larger than in previous studies, and clearly had sufficient power to detect typical Ganong effects, as the lexicality-based Ganong effect was highly significant. These results suggest that, at least in some cases, frequency does not exert the same influence on phonological judgment as lexicality does.

Could the results be due to something special about Chinese? There is psycholinguistic evidence that segments play a less important role in Chinese word recognition than they do in other languages (e.g., O'Seaghdha *et al.*, 2010), but it is not clear why a segmental contrast would yield a lexicality Ganong effect but not a frequency one.

Another potential explanation could be that the Ganong effect is a result of controlled processes that depend on participants' awareness, and that participants are not aware of a frequency difference the way they are aware of a lexicality difference. However, this explanation would be inconsistent with recent convincing evidence that Ganong effects emerge quickly and are probably not dependent on late decision-making processes (e.g., Kingston *et al.*, 2016; Rysling *et al.*, 2015).

Another possibility, suggested by an anonymous reviewer, is that the stimuli in the present study were mostly from the low end of the VOT range for Mandarin, which could lead to biases. However, the stimuli in the present study may have a naturally lower VOT range than previously published norms, given that they come from a different speaker and are produced in a different context. We normed our stimuli with a categorical discrimination task and chose continua centered on the categorical boundary for each continuum. Thus, we assume the stimuli were not biased, but were sampled equally from the "aspirated" and "unaspirated" sides of the continuum for this particular speaker with these particular items in this particular recording context. Furthermore, it is not clear what mechanism would cause this sort of distributional bias to override potential frequency-based Ganong effects but not lexicality-based Ganong effects.

Why did frequency-based Ganong effects emerge in two previous studies and not this one? We cannot rule out the possibility that the conclusion of the present study simply represents a type II error, or the previous studies a type I error; more studies are needed to provide sufficient data to rule these possibilities out. Barring those explanations, is there any systematic moderator that could explain why frequency-based Ganong effects seem to occur in some studies (or with some stimuli) and not others? Things that look like Ganong effects may emerge for lower-level reasons like differences in phoneme transition probability (Norris *et al.*, 2000) or neighbourhood density of continuum endpoints (Newman *et al.*, 1997). It is not likely that this could account for Connine and colleagues' (1993) findings, as they used a large number of items and transitional probability could only account for their effects if it were confounded with frequency in all or most of their items. Transitional probability or some other low-level factor like it, however, may account for the apparent Ganong effect observed in Shen and Politzer-Ahles (2018)—or the lack of effect in the present study. Attempting to tease apart these effects from frequency is a valuable direction for future study, and may provide some insight on the apparent variability in frequency-based Ganong effects. We note, however, that the operationalization of these psycho-phonological constructs is a nontrivial problem, particularly in a language like Mandarin (see, e.g., Sharma and Yao, 2017; Neergard *et al.*, 2016) but really in all languages, since neighbourhood measures are highly dependent on how

one defines a neighbour and on one's theory of the language's phonological structure (Turnbull, 2019). While one of these factors could account for the present results, we do not think it is very likely; it would require that neighbourhood density (or transitional probability or some other such factor) be confounded with aspirated–unaspirated bias in the same way across all three pairs of continua we tested for the frequency Ganong effect, and in the opposite way as that across all three pairs of continua we tested for the lexicality Ganong effect.

Finally, another potential explanation is that the difference in frequency between continuum endpoints was too small in some continua, or even in the wrong direction (see Table 1 for a discussion). This was our motivation for using multiple continua, rather than relying on a single continuum [as in our previous study (Shen and Politzer-Ahles, 2018)]; even if one continuum is arguably problematic depending on the choice or corpus or the cutoff for how much different the frequencies need to be to observe an effect, hopefully other continua would be able to show the effect. However, as shown in Sec. 3, none of the continua showed a significant Ganong effect in the expected direction, which suggests that the lack of effect was not due to problems in any particular item/continuum.

Overall, we do not yet have a satisfactory explanation for why frequency-based Ganong effects seem to occur in some situations and not others. The present study reveals a gap in our understanding of Ganong effects, and demonstrates the need for further research to better elucidate these problems.

Acknowledgments

References and links

Baayen, R., Davidson, D., and Bates, D. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Memory Lang. **59**, 390–412.

Barr, D., Levy, R., Scheepers, C., and Tily, H. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Memory Lang. **68**, 255–278.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**, 1–48.

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer" [Computer program]. Version 6.0.30, http://www.praat.org/ (Last viewed July 22, 2017).

Cai, Q., and Brysbaert, M. (**2010**). "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles," PLoS One **5**, e10729.

Chen, K., Huang, C., Chang, L., and Hsu, H. (**1996**). "SINICA CORPUS: Design methodology for balanced corpora," in *11th Pacific Asia Conference on Language, Information and Computation*, pp. 167–176.

Connine, C., Titone, D., and Wang, J. (**1993**). "Auditory word recognition: Extrinsic and intrinsic effects of word frequency," J. Exp. Psychol. **19**, 81–94.

Forster, K., and Forster, J. (**2003**). "DMDX: A Windows display program with millisecond accuracy," Behav. Res. Methods, Instrum. Comput. **35**, 116–124.

Ganong, W. (**1980**). "Phonetic categorization in auditory word perception," J. Exp. Psychol. **6**, 110–125.

Kingston, J., Levy, J., Rysling, A., and Staub, A. (**2016**). "Eye movement evidence for an immediate Ganong effect," J. Exp. Psychol. **42**, 1969–1988.

Neergard, K., Xu, H., and Huang, C. (**2016**). "Database of Mandarin neighborhood statistics," in *Proceedings of Language Resources and Evaluation*.

Newman, R., Sawusch, J., and Luce, P. (**1997**). "Lexical neighbourhood effects in phonetic processing," J. Exp. Psychol. **23**, 873–889.

Norris, D., McQueen, J., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–370.

O'Seaghdha, P. G., Chen, J.-Y., and Chen, T.-M. (**2010**). "Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English," Cognition **115**, 282–302.

R Core Team (**2016**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (Last viewed January 6, 2020).

Rysling, A., Kingston, J., Staub, A., Cohen, A., and Starns, J. (**2015**). "Early Ganong effects," in *Proceedings of the International Congress of Phonetic Sciences*.

Sharma, B., and Yao, Y. (**2017**). "What is in the neighbourhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese," LSA Extended Abstracts.

Shen, L., and Politzer-Ahles, S. (**2018**). "Analysis of the influence of word frequency in auditory perception," Poster presented at Hanyang International Symposium on Phonetics and Cognitive Sciences of Language.

Turnbull, R. (**2019**). "Choices in abstract phonological analysis have direct consequences for psycholinguistic predictions: The case of phonological neighbourhood networks," Poster presented at Hanyang International Symposium on Phonetics and Cognitive Sciences of Language.