Top-Down Biases for Lexicality and Frequency in Both Monosyllabic and Disyllabic Stimuli: Evidence from Cantonese

Ka Keung Lee (ka_keung_lee@sfu.ca)
Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

Stephen Politzer-Ahles (sjpa@ku.edu)
Department of Linguistics, the University of Kansas, 420W Blake Hall Lawrence, 1541 Lilac Ln, Lawrence, KS 66045, United States

Abstract

Also known as the Ganong effect, a lexicality bias effect- i.e., the bias to interpret an ambiguous sound as the phoneme that yields a real word in its context - has been widely replicated. The search for a similar frequency bias effect, on the other hand, has yielded mixed results: In English, a bias has been observed such that listeners tend to interpret an ambiguous sound as the phoneme that yields a higher-frequency word, but this has failed to be replicated in Mandarin. One difference between these studies is the use of monosyllabic vs. disyllabic stimuli. To determine the factors that influence the presence of a bias effect, the present study tested for both frequency and lexicality bias effects using monosyllabic and disyllabic stimuli in Cantonese. Results show that the lexicality and frequency bias effects can be elicited in both monosyllabic and disyllabic stimuli, but the frequency effect is weaker.

Keywords: Cantonese; Ganong effect; lexical activation; top-down processing; speech perception.

Introduction

When English speakers hear a sound that is ambiguous between, for instance, /t/ and /d/, in the context of "_*ask*", they tend to judge the sound they heard as /t/ instead of /d/, presumably because /t/ yields a real word "task" whereas /d/ a pseudoword "dask" (Ganong, 1980). This phenomenon is called the "Ganong effect".

Using English words pairs started with either voiced (/b/, /d/, /g/) or voiceless (/p/, /t/, /k/) onset while manipulating their frequency rate, Connine and colleagues (1993) found a similar preference for judgments that yield a higher-frequency word, as opposed to a lower-frequency word. For example, when English speakers were presented with a stimulus that was ambiguous between "time" and "dime", they tended to judge it as "time", presumably because "time" is used more frequently. Based on these results, it has been argued that speech recognition shows a frequency bias effect (we noted that the term "bias effect" was used elsewhere to refer specifically to post-access decision processes, we are not using it in this sense or implying a commitment to any particular model of what causes this effect, but to refer to listeners' tendency to judge a sound as whichever phoneme yields a higher-frequency word) similar to the abovementioned lexicality bias effect (i.e., Ganong effect).

While the bias effect for lexicality has been widely replicated in different languages, the one for frequency is less understood. Since the original report of the frequency bias effect by Connine and colleagues (1993), there have been few attempts to replicate it. A bias effect experiment on Mandarin that manipulated the lexical tone by Yang and colleagues (2019) found evidence consistent with a frequency bias effect (although the study was not specifically designed to test this). In contrast, another Mandarin study that manipulated voice onset time (VOT) by Politzer-Ahles and colleagues (2020) failed to find such an effect – in the latter study, the extent to which an ambiguous phoneme was interpreted as /t/ (for example) was not affected by whether it would yield a higher- or lower-frequency word. Thus far, there is no satisfactory explanation for why the frequency bias effect has occurred in some studies and not others.

We note that most studies that tested the lexicality or frequency bias effects used monosyllabic stimuli (Connine et al., 1993; Fox & Unkefer, 1985; Ganong, 1980; Kingston et al., 2016; Yang et al., 2019; Yang et al., 2022), whereas Politzer-Ahles and colleagues (2020), who found the lexicality bias effect but not the frequency bias effect, used disyllabic stimuli (a pilot study by Shen and colleagues [2018] found a frequency bias effect with disyllables but only used one item per condition, so that effect may have been due to some factor other than frequency). This difference in stimulus length may cause the latter study not to find a frequency bias effect.

There is independent evidence that word length affects lexical activation; for example, Pitt and Samuel (2006) found that longer words (e.g., trisyllables) would generate stronger lexical activation than shorter words (e.g., monosyllables) because longer words, compared to shorter ones, can provide more bottom-up evidence and relatively less inhibition triggered by competition from similar words. Their results remained the same even when they compressed the longer words and expanded the shorter words to last for the same duration; thus, the effect is not based on the amount of processing time available to participants, but on the amount of information (e.g., number of phonemes) provided by the stimuli. Therefore, the finding of more lexical activation in longer words may explain why Politzer-Ahles and colleagues' (2020) disyllables elicited a lexicality bias effect but not a frequency one.

As for lexicality, the bias effect should still be detectable even with longer disyllabic stimuli. When hearing an ambiguous phoneme that would render the context either a real word or not, there would be lexical activation only for

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

the real word and none for the pseudoword; therefore, listeners would be biased towards the phoneme interpretation that yields the real word, regardless of whether the carrier stimulus is short or long.

On the other hand, when hearing an ambiguous phoneme that would render the stimulus either a higher-frequency or a lower-frequency word, the outcome may depend on how long the word is. When the target word is short (e.g., a monosyllabic word like "time"), there is little bottom-up information available to aid word recognition. With less bottom-up information available, top-down influences, such as a frequency bias effect, may play a significant role. It is possible that by the time the listeners are required to respond, only the higher-frequency word has been activated. Therefore, the speakers would prefer to choose the option that would yield a higher-frequency word.

The outcome would be different, however, when the ambiguous phoneme occurs within a longer word stimulus with more bottom-up information, e.g., a stimulus that is ambiguous between "tessellation" (/t̲esəleɪʃn/) and "desolation" (/d̲esəleɪʃn/). Although "desolation" may be more frequent than "tessellation", both might be long enough that acoustic realizations of these words have provided substantial bottom-up support for the respective lexical representations. If that is the case, we might observe a ceiling effect. Since the stimulus is relatively long, both the higher-frequency word "desolation" and the lower-frequency word "tessellation" receive so much activation that they are both highly activated by the time the listeners need to make a judgment. In this case, lexical activation may be strong enough that top-down influences like the frequency bias effect no longer have a role to play, or that the role they can play is so negligible that it is not detectable in typical experiments.

To investigate whether the frequency bias effect is moderated by syllabic numbers, the present study used Cantonese monosyllabic and disyllabic stimuli in a Ganong effect paradigm (Ganong, 1980). As a tonal language with more than 85 million speakers (Eberhard et al., 2025), Cantonese consists of a larger monosyllabic lexical inventory (e.g., Alderete et al., 2017) and has not been well-documented in the literature (e.g., Tucker & Wright, 2020) compared to Mandarin, and the lexicality and frequency bias effects have not been previously studied in Cantonese to our knowledge. Therefore, conducting a study to explore whether the lexicality and frequency bias effects can be replicated in Cantonese will both fill a gap in the literature and shed new light on the less well-understood frequency bias effect.

Method

Experimental procedures were approved by the Human Subjects Ethics Sub-Committee at the Hong Kong Polytechnic University. The main test list, the training stimuli list, and all data and stimuli are available at https://osf.io/6cbgy/.

Participants

Fifty native speakers of Hong Kong Cantonese took part in the experiment at the Hong Kong Polytechnic University. Two participants were removed because they were not born in Hong Kong and one participant's data was lost due to a computer crash during the task; this left 47 participants in the final dataset (average age 21.9 years, range 18-29, 19 men and 28 women). All 47 participants self-reported that they were born in Hong Kong, were not diagnosed with any hearing or reading disability, and were not majoring in linguistics or psychology. Apart from Cantonese, the 47 participants all reported they could also speak English, 36 could speak Mandarin, 7 could speak Japanese, 3 could speak Korean, and 1 each could speak French, Hakka, Indonesian, Teochew, and Spanish. All participants provided written informed consent and were compensated HK$100 for their participation.

Stimuli

Four sets of monosyllabic and three sets of disyllabic alveolar-stop-initial stimuli were selected to form four contexts, namely contexts (Cont.) in which the aspirated stop /t/ yields real words (*t* lexi. bias), pseudowords (*d* lexi. bias), higher-frequency (*t* freq. bias) and lower-frequency (*d* freq. bias) words (16 monosyllables: 4 sets * 4 conditions; 12 disyllables: 3 sets * 4 conditions; see Table 1).

Table 1: Stimuli Examples. [1]

Cont.	Monosyllable		Disyllable	
	Stim.	Freq.	Stim.	Freq.
t lexi. bias	*{t/d}ang4* 'to fly' vs. gap	103 (160) vs. gap	*{t/d}ai2jim6* 'to experience' vs. gap	42 vs. gap
d lexi. bias	*{t/d}ang1* gap vs. 'to mount'	gap vs. 1,855 (2,378)	*{t/d}ai2sin3* gap vs. 'the bottom line'	gap vs. 53
t freq. bias	*{t/d}ai2* 'to look' vs. 'the bottom'	5,786 (7,176) vs. 2,140 (3,660)	*{t/d}ai2hei3* 'to watch a movie' vs. 'the mental strength'	54 vs. 1
d freq. bias	*{t/d}ai1* 'the ladder' vs. 'low'	172 (172) vs. 1,701 (2,624)	*{t/d}ai2sei2* 'to look down on' vs. 'to deserve a bad consequence'	12 vs. 17

Regarding the four sets of monosyllabic stimuli, the rhyme (nucleus and coda) and the tone in each contrasting pair are identical, leaving the VOT as the only difference within each pair. For the 3-set disyllabic stimuli, the rhyme (nucleus and coda) and the tone of the first syllable and the whole second syllable structure in each contrasting pair are identical, only the VOT of the first syllable differs. The frequency data were

---

[1] (A) The **Stim.** (stimuli) column lists the *Jyutping* transcription of each stimulus and its respective meanings. (B) In the **Freq.** (Frequency) column, the number outside the bracket indicates the frequency rank of the most frequently used characters/words under the corresponding *Jyutping* transcription; the one inside the bracket is all possible homophones' frequency ranks added up. The "gap" means the *Jyutping* transcription in question, with the respective marked tone, cannot form a real character/word in Cantonese.

collected from the *Jyutdin* database (Jyutdin team, 2020), the latest update when this measure was collected was on October 17th 2020. The discourses in the *Jyutdin* database were collected from daily conversations of the database members and the composed articles from the first (in 2014) and the second (in 2016) *Canton Compo competition* hosted by *Societas Linguistica Hongkongensis*. The database consists of two frequency data lists: Character Usage Frequency List (monosyllable) and Word Frequency List (non-monosyllable), containing 2,648,243 characters and 2,927,707 words respectively.

In Cantonese, a single phonological form often corresponds to various characters that possess different semantic meanings and orthographic forms. For instance, Cantonese *he3* can mean ‘air’, ‘a movie’, ‘a vessel’, or ‘to abandon’ with the orthographic forms “氣”, “戲”, “器”, or “棄” respectively. In addition, Hong Kong Cantonese speakers might confuse tone 2 and tone 5, tone 3 and tone 6, and even tone 4 and tone 6 (tone mergers; Mok et al., 2010). For instance, people might perceive *daai3* ‘a belt’ as *daai6* ‘big’. In this case, it would be a challenge for us to design our target stimuli, especially when we try to decide whether the selected monosyllable would render a higher- or a lower-frequency word.

For example, suppose we chose “*{t/d}aa3*” as one of our stimuli, and let’s say we found two homophones both pronounced *taa3* with frequencies of 10 and 2 respectively, and four homophones pronounced *daa3* with frequencies of 7, 6, 6, and 5 respectively. If we only considered the most frequent homophone, *taa3* (frequency of 10) would be more frequent than *daa3* (frequency of 7). But if we added up all possible homophones’ frequencies, then *daa3* (7+6+6+5=24) would be more frequent than *taa3* (10+2=12). This example illustrates the challenge caused by homophony.

Another challenge came from tone mergers. Imagine we also found one character pronounced as *taa6* with a frequency of 4, and another character pronounced as *daa6* with a frequency of 1. So far, *taa6* appears to be more frequent than *daa6*. Recall, though, that many listeners might find it difficult to distinguish *taa6* from *taa3* or *daa6* from *daa3*, so a person hearing *taa6* or *daa6* might also activate forms associated with *taa3* or *daa3*. If that is the case, then a more accurate measure of the frequency of any of these sounds would be the sum of the tone 6 and tone 3 counterparts. In this case, *taa6* would actually be less frequent (frequency of 4 for *taa6* plus frequency of 12 for *taa3*, for a total frequency of 16) than *daa6* (frequency of 1 for *daa6* plus frequency of 24 for *daa3*, for a total frequency of 25).

Therefore, when selecting stimuli for the frequency-biased conditions, we ensured that the higher-frequency end of the continuum fulfilled the following three criteria (taking *taa3* as an example here): 1) The most frequently used character under *taa3* has to be more frequent than the most frequently used *daa3* character; 2) the total frequency that all *taa3* homophones added up has to be higher than the one that all *daa3* homophones added up; 3) due to the tone mergers, the total frequency of all homophones under *taa3* and *taa6* (merge tone 3 and tone 6) has also to be higher than the one under *daa3* and *daa6* if that were the case.

Regarding the real word and pseudoword pairs, for instance, if it were the case that *taa3* and *daa6* were real words but *daa3* a pseudoword, then, due to the tone mergers, participants might also perceive *daa3* to be a real word too (mistake *daa3* for *daa6*). Therefore, when we selected characters that are in tone 2, 3, or 4 as pseudowords, we also ensured that they are also pseudowords when they are in tone 5 (merges with tone 2) or 6 (merges with tone 3 and tone 4).

According to the Character List and Character Usage Frequency List in the *Jyutdin* database (Jyutdin team, 2020), both *tin3* and *tim3* have 0 occurrence, although they can be possible characters in Cantonese (*tin3*: 琠, 瑱; and *tim3*: 掭, 䑙). In other words, those characters are either rarely or never used in Hong Kong Cantonese. Therefore, the present study considers *tin3* and *tim3* to be both pseudowords. A post-experiment survey was also conducted to ensure that *tin3* and *tim3* were not real words for the participants.

All stimuli in the present study were recorded in a sound-attenuated booth (at a sampling rate of 44.1 kHz) by a male native speaker of Cantonese (aged 26). The onset was manipulated and split into an unaspirated-to-aspirated continuum (10 to 60 ms of VOT, in 5-ms steps; the shorter the VOT, the more the stimulus sounds like /d/) using Praat (Boersma & Weenink, 2017). In the experiment, we used an eleven-step continuum centered on the categorical boundary (around 35 ms) with 5-ms increments between steps, in the end, yielding 308 tokens (11 steps * 28 [16 monosyllables + 12 disyllables] stimuli) in total.

All stimuli were delivered using DMDX (Forster & Forster, 2003). Each trial began with a fixation point marked with an “x” symbol indicating the start of a trial. The prompt showed two given options: The aspirated /t/ option (the printed letter “T”) was always on the left of the monitor, and the unaspirated /d/ option (the printed letter “D”) was always on the right. The prompt appeared on the screen at the same time the stimulus played. The participants were instructed to decide whether the initial segment (onset) of the stimulus they heard is /t/ or /d/ by pressing the left or right Shift buttons, respectively, on a computer keyboard as soon as possible. The participants had 8 seconds to make their responses before the program proceeded to the next trial.

Procedure

The experiment contained three sections: 1) the monosyllabic and 2) disyllabic blocks, and 3) the post-experiment survey. All written instructions were in traditional Chinese. The whole experiment took around 60 minutes.

Monosyllabic and disyllabic stimuli were presented in separate blocks, with block order counterbalanced across participants. Each stimulus type began with a practice block and then multiple blocks of critical stimuli with breaks in between. The monosyllabic stimuli were presented over 8 blocks (preceded by the monosyllabic practice block) and the disyllabic stimuli over 6 blocks (preceded by the disyllabic practice block). Half of the participants completed the monosyllabic blocks first, and the other half completed the disyllabic blocks first. During the practice block, the participants were seated in front of a computer monitor and listened to the stimuli at a comfortable volume through headphones in an individual soundproof computer room. Each practice block included 8 items with unambiguous VOTs. Before the training section, all participants would be instructed to decide whether the onset they heard is “T” or “D” in the monosyllabic section; and whether the onset of the first syllable they heard is “T” or “D” in the disyllabic section by pressing either the left or right Shift button. Although most Cantonese speakers in Hong Kong did not receive school-based training for writing Cantonese in an alphabetic orthography, they are generally familiar with the use of alphabetic letters to correspond to the Cantonese /t/ and /d/ phonemes (e.g., they know *tai1* ‘the ladder’ starts with /t/ and *dai1* ‘to be low’ starts with /d/) and they also use Romanized letters in texting in everyday life. Before the main test, the experimenter in charge also made sure that the participants understood the requirements and procedure of the task. Practice materials were not reused in the main test.

The procedure of the main test was identical to the training section except that the stimuli included those with ambiguous VOTs. In each given block, all 44 tokens of a four-context target stimulus (11 steps * 4 contexts) were played once in random order, and then they were repeated one more time in random order, such that there were altogether 88 trials (44 target tokens * 2 repetitions) in each block. Each set of stimuli occurred in two blocks; therefore, each token was responded to four times (2 repetitions * 2 blocks). Each participant completed 1,232 trials (11 steps * 28 [16 monosyllables + 12 disyllables] targets * 4 repetitions) in the main test and was given self-paced breaks between blocks.

After the main test, all participants filled out a two-section survey form. In the first section, the participants judged which characters or words are more frequently used in their daily lives. The characters and words (in audio format) in the first section were the stimuli from the main test and were presented to the participants auditorily with unambiguous VOTs (these stimuli were the unedited natural productions). In the second section, the audio files of *tin3* and *tim3* were played, and the participants indicated whether they knew any possible character with these two pronunciations; this was done to confirm that these two characters are pseudowords for the participants.

Results

Trials that timed out at 8,000 ms were excluded from the analysis. Apart from that, trials with reaction times below 200 ms were excluded based on the assumption that the responses that fast are likely to be mistakes, given that it typically takes more than 200 ms to prepare and execute a response to words in isolation (200 ms is a typical cutoff in lexical decision studies, for instance). This trimming resulted in a loss of 0.2% of the data.

In the post-experiment survey, the majority of our participants reported that they did not know any character pronounced as *tim3* (43 out of 47) and *tin3* (46 out of 47), and their responses also matched the frequency rates that we collected from the *Jyutdin* database (Jyutdin team, 2020) except the *{t/d}ing1* pair (analysis with or without this item excluded did not change the results).

Figure 1 shows the results of the experiment: The left figure shows the results from monosyllabic stimuli, and the right figure shows the results from disyllabic stimuli. The blue lines represent the proportion of /t/ judgments in stimuli that are biased towards /t/[2] (solid lines for lexicality and dotted lines for frequency); thus, bias effects occurred wherever the blue line is above the corresponding red line.

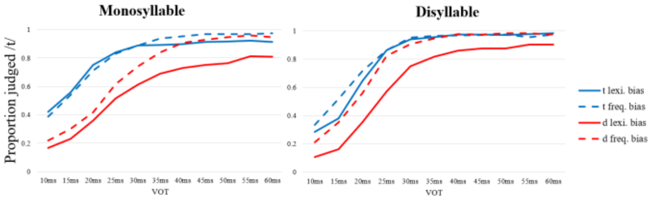

Figure 1: Lexicality and frequency bias effect in disyllabic and monosyllabic stimuli.

Figure 2 shows the comparisons of the lexicality and frequency bias effects in the monosyllabic and disyllabic stimuli: The first row illustrates the comparison of the lexicality (solid lines) and frequency (dotted lines) bias effects within monosyllabic (orange lines) and disyllabic (green lines) stimuli respectively, and the second row illustrates the comparison of the influence of syllable number on the lexicality and frequency bias effects respectively. Based on the figure, it seems that the lexicality bias effect is more robust than the frequency bias effect in both monosyllabic and disyllabic stimuli. These two types of bias effects appear to be stronger in the monosyllabic stimuli than in the disyllabic stimuli.

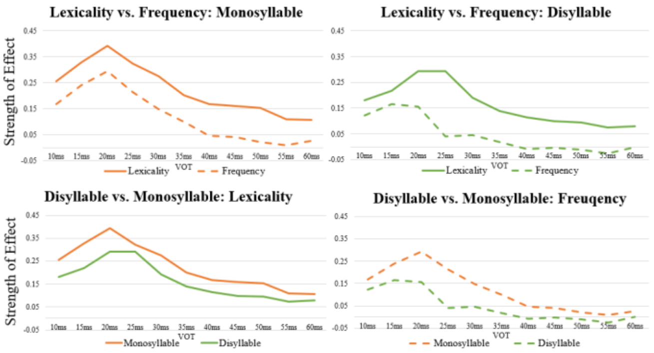

Figure 2: Bias effects comparison.

---

[2] The percentage of the proportion judged aspirated (the value plotted on the y-axis) was calculated by using the proportion of /t/ responses in *t* bias contexts minus the proportion of /t/ responses in *d* bias contexts.

Figure 3 shows the results of the individual participants: The first row displays the individual lexicality (solid lines) and frequency bias effects (dotted lines) from monosyllabic stimuli respectively, and the second row displays the individual lexicality and frequency bias effects from disyllabic stimuli respectively. The orange and green lines represent the respective average monosyllabic and disyllabic lexicality and frequency bias effects of all participants. From Figures 1 to 3, it is apparent that both lexicality and frequency bias effects occurred in both monosyllables and disyllables, but that the frequency effects appear to be smaller than the lexicality effects, and the effects in disyllables appear to be smaller than the effects in monosyllables. Statistical tests of these observations are described below.

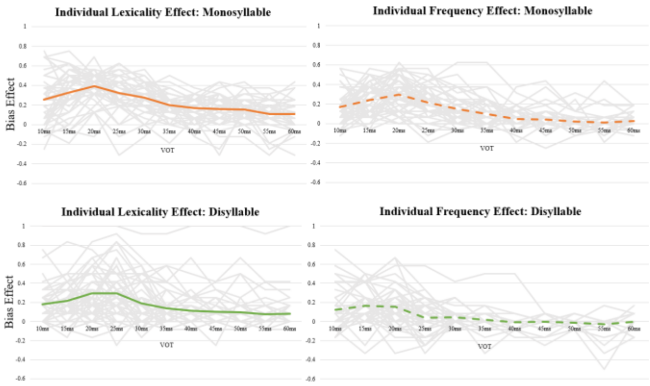

Figure 3: Results of the individual bias effects.

We performed two sets of statistical analyses: One to examine whether a bias effect occurred at all in each condition, and another to compare the sizes of the bias effects between various conditions. By “condition” here, we mean each combination of bias type and the number of syllables, i.e., each cell in Figure 3.

To test for the presence of a bias effect in each condition, we ran mixed-effects logistic models (Baayen et al., 2018) on each condition separately. Models were implemented in the lme4 package (Bates et al., 2015) in R (R core team, 2022). For each condition, we regressed responses (with D responses coded as 0 and T responses coded as 1, such that positive coefficients represent more /t/ responses) on the fixed effects Bias (whether the continuum was lexically biased towards /t/ or /d/), continuum Step, and their interaction. The random effect structure included the same parameters as the fixed effects (intercept, Bias, Step, and the interaction between Bias and Step) for each participant, but did not include correlations between random effects (models including this -- the maximal random effects structure -- did not successfully converge). The full model was “*Response ~ Bias * Step + (0 + Bias * Step || Participant)*”[3]. Bias was dummy-coded with /d/-biased continua as 0 and /t/-biased continua as 1, such that positive coefficients represent bias effects, i.e., more T responses in /t/-biased continua. Step was centered. We are primarily interested in the main effect of Bias, which represents bias effects (i.e., a shift in the categorization function); significant *Step*Bias* interactions would represent steeper slopes for one continuum than another, which we do not have any a priori predictions about. One-tailed tests were used in this analysis, given that only effects in one direction (i.e., effects consistent with the bias) could be taken as bias effects, and we had no interest in testing for effects in the opposite direction.

Monosyllabic continua had highly significant effects for both lexicality ($b = 2.53$, $z = 10.79$, $p < .001$) and frequency ($b = 1.11$, $z = 11.66$, $p < .001$). In disyllables, the lexicality bias effect was also significant ($b = 2.21$, $z = 8.93$, $p < .001$), as well as the frequency bias effect ($b = 0.30$, $z = 1.94$, $p = .026$).

To compare the size of the bias effect in different conditions, we compared bias effects within the same model by letting the Bias effect interact with the number of syllables or with bias type (a separate model was run for each of these tests because a full model with all possible interactions did not converge, because of the low number of observations per cell). Bias effects in disyllables were numerically smaller than in monosyllables, but not significantly so, as indicated by the non-significant interaction between bias and the number of syllables ($b = -0.25$, $z = -1.49$, $p = .137$) in a model including these two variables as well as Step and all interactions between these variables.[4] The frequency bias effect was significantly smaller than the lexicality bias effect, as indicated by a significant interaction between bias and type of bias (lexicality bias vs. frequency bias; $b = -0.77$, $z = -7.40$, $p < .001$).

While the bias effect in disyllables was not significantly smaller than that in monosyllables when analyzing the whole dataset, Figures 1 to 3 suggest that it is smaller, particularly within the frequency-biased conditions. This pattern, however, was not significant (three-way interaction between bias, bias type, and number of syllables: $b = 0.29$, $z = 1.17$, $p = .242$; this is based on a model with bias, bias type, number of syllables, and all interactions between them, as well as a coefficient for Step not interacting with the other variables: *Response ~ BiasType * NSyllables * Bias + Step + (BiasType * NSyllables * Bias + Step || Participant*). A full model [including all possible interactions with Step] did not converge because of the small amount of data in each cell and the large number of zero cells. Thus, while the frequency bias in disyllables appears to have a smaller effect than any other condition, there was no statistical support for this pattern.

---

[3] We removed the random intercept here because the model including a random intercept did not converge. During the model simplification, we would first take out a random intercept before removing a random slope to keep the random effect in the model.

[4] In an exploratory analysis examining the effect of the number of syllables within the lexicality and frequency conditions separately, disyllables elicited both significantly smaller lexicality bias effects [$b = -0.75$, $z = -3.12$, $p = .002$] and frequency bias effects [$b = -0.67$, $z = -5.26$, $p < .001$] than monosyllables did.

Discussion

The present study observes the lexicality and frequency bias effects in both monosyllables and disyllables when the VOTs of the stimuli were manipulated. The results demonstrate that both types of aforementioned bias effects can be replicated in Cantonese, a tonal language that does not belong to the Indo-European family.

In general, our results are in line with the previous studies that found larger bias effects (Connine et al., 1993; Fox & Unkefer, 1985; Ganong, 1980; Kingston et al., 2016; Politzer-Ahles et al., 2020; Shen & Politzer-Ahles, 2018) at the intermediate VOTs (around 20 ms to 35 ms). Notwithstanding, it is still hard for participants to change their responses when the VOT is clearly unaspirated (e.g., 10ms) or clearly aspirated (e.g., 60ms), ambiguous sounds (intermediate VOTs) enable the top-down bias effect to play a role during a decision task. A similar pattern is observed for both lexicality and frequency bias effects, suggesting they might belong to a similar speech recognition processing, but the numerically weaker frequency bias effect might be the result of the competition between the activated higher- and lower-frequency words cancelling (a portion of) the effect.

A possible reason that we replicated the frequency bias effect but Politzer-Ahles and colleagues (2020) did not might be that Mandarin speakers and Cantonese speakers access the mental lexicon differently, since Mandarin speakers have received formal training in phonemic transcription at school, but not for Cantonese speakers. It is not clear, however, why this would cause Cantonese speakers to have a different bias effect than Mandarin speakers, as opposed to having an across-the-board higher or lower sensitivity to VOT differences regardless of lexicality. Furthermore, if the difference between the present results and the previous Mandarin results were because Mandarin speakers have trained with an alphabetic writing system and Cantonese speakers have not, then we would expect Mandarin and English speakers to pattern together, contra the real pattern of results (in which the English speakers of Connine and colleagues [1993], and the Cantonese speakers of the present study pattern together, separate from the Mandarin speakers of Politzer-Ahles and colleagues).

One limitation of the current study is that most people in Hong Kong are bilingual or trilingual, typically speaking Cantonese along with English and/or Mandarin. As a result, we were unable to recruit only monolingual participants. Since the influence of additional language knowledge on speech perception is still unclear, we cannot rule out the possibility that exposure to other languages affected participants' top-down processing. For instance, knowledge of another language may shift the categorical boundary between /t/ and /d/ (e.g., Yang, 2021), potentially influencing participants' perception. Additionally, due to the lack of a phonotactic probability database for Cantonese, future studies could replicate our experiment in a language like English to examine whether phonotactic probability influences the size of the bias effect.

Overall, the present study reveals that the bottom-up lexical activation may play an important role in determining whether a top-down frequency bias effect is observable in a study that requires participants to make a conscious decision. The syllable number of the stimuli may be a crucial factor that affects the level of lexical activation; while shorter words will receive less activation, longer words will be activated more due to the lesser competition from similar words. Since a pseudoword presumably will not receive any activation to compete against a real word, a lexicality bias effect is easier to observe regardless of the syllable number of the stimuli. A frequency bias effect may not be too difficult to detect in an experiment using shorter words (e.g., monosyllables) because lexical activation of the lower-frequency word is small and may not be enough to cancel the frequency bias effect. But when the target word becomes longer (e.g., disyllable), the lower-frequency word may be activated more, thus attenuating (a portion of) the frequency bias effect. In this case, the frequency bias effect may be harder to observe.

References

Alderete, J., Chan, Q., Chan, M., Fan, G. & Nickel, O. (2017). *Cantonese grammar synopsis*. Department of Linguistics, Simon Fraser University.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2018). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412.

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

Boersma, P. & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.30). http://www.praat.org.

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

Eberhard, D. M., Gary, F. S. & Charles, D. F. (2025, January 27). *Ethnologue: Languages of the World-Yue Chinese* https://www.ethnologue.com/language/yue/.

Forster, K. I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers,* 35, 116-124.

Fox, R. A. & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, 13, 69-90.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental psychology: Human perception and performance, 6*(1), 110-125.

Jyutdin team. (2020, October 17). *Jyutdin database*. https://words.hk/faiman/analysis/.

Kingston, J., Levy, J., Rysling, A. & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. Journal of experimental psychology: *Human perception and performance, 42*(12), 1969-1988.

Mok, P. K. & Wong, W. Y. (2010). Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. *Proceedings of Speech Prosody 2010 Fifth International Conference, 100916*, 1-4. http://ling.cuhk.edu.hk/people/peggy/SP2010_perception.pdf

Pitt, M. A. & Samuel, A. G. (2006). Word length and lexical activation: longer is better. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1120-1135.

Politzer-Ahles, S., Lee, K. K. & Shen, L. (2020). Ganong effects for frequency may not be robust. *The Journal of the Acoustical Society of America*. 147, EL37-EL42.

R core team. (2022). *R: A language and environment for statistical computing* (Version 4.2.1). https://www.R-project.org/.

Shen, L. & Politzer-Ahles, S. (2018, May 18-19). *Analysis of the influence of word frequency in auditory perception*. Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2018, Seoul, South Korea.

Tucker, B. V. & Wright, R. (2020). Introduction to the special issue on the phonetics of under-documented languages. *The Journal of the Acoustical Society of America, 147*(4), 2741-2744.

Yang, J. (2021). Comparison of VOTs in Mandarin–English bilingual children and corresponding monolingual children and adults. *Second Language Research, 37*(1), 3-26.

Yang, T. H., Jin, S. J. & Lu, Y. A. (2019). The effect of Mandarin accidental gaps on perceptual categorization. *Proceedings of the 19th International Congress of Phonetic Sciences*. 2022-2026. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2071.pdf

Yang, T. H., Jin, S. J. & Lu, Y. A. (2020). The effect of lexicality, frequency, and markedness on Mandarin tonal categorization. *Frontiers in Psychology*, 13, 4190-4197.