

# Gradient Acceptability in Mandarin Nonword Judgment

Shuxiao Gong & Jie Zhang  
University of Kansas

# Theoretical Background

- Speakers offer gradient judgment on novel words (e.g. blick > bwick > bnick in English), which is guided by their phonotactic knowledge (Berent et al. 2007, Hayes & White 2013).
- Such gradient acceptability is reported mainly in Indo-European languages like English with complex phonotactics.
- Mandarin has simpler phonotactics, and Mandarin syllables are processed as a whole (O'Seaghdha et al. 2010).
- It is worthwhile to explore whether phonotactic acceptability is still gradient in languages like Mandarin.

# Lexical Statistics

- One of the sources of this gradient judgment is the statistical properties of the lexicon. How similar a non-word is compared to existing words.
  - Phonotactic probability (Vitevitch & Luce 2004)
  - Neighborhood density (Bailey & Hahn 2001)

# Lexical Statistics is not Sufficient

- Berent et al. (2007) tested the perception of three types of English non-words with unattested onset clusters.
- These missing onset clusters vary in terms of sonority profile: rising (e.g., **bn**), level (e.g., **bd**), or falling (e.g., **lb**).
- Three types of non-words were controlled for lexical statistics.
- Results: non-words with falling sonority onsets (**lb**) were perceived least accurately, followed by level (**bd**), and rising (**bn**).

# Lexical Statistics is not Sufficient

- Coetzee (2008) compared the acceptability judgment of English non-words [spVp] and [skVk].
- The grammar predicts that [skVk] is more well-formed than [spVp], because constraints on labial co-occurrence are more common than dorsal.
- The stimuli were constructed such that grammaticality and lexical statistics conflict: [spVp] forms have higher statistical measures than [skVk] forms.
- Results: [spVp] forms were still rated lower than [skVk].

# Grammaticality in Non-word Judgment

- These non-word judgment experiments indicate that there are universal preferences in speech perception, such as Sonority Sequencing Principle and Obligatory Contour Principle.
- These preferences constitute arguments for the independent effect of grammaticality.
- The gradient non-word acceptability, thus, is the result of the interaction between grammaticality and lexical statistics (Shademan 2007, Hayes & White 2013).

# Grammatical Factors of This Study

- Non-words violating some principled phonotactic constraints should be labelled as systematic gaps, whereas non-words that do not violate those constraints are accidental gaps (Chomsky & Halle 1965).
- The gradient acceptability implied:  
Real word > Accidental gap > Systematic gap

# Systematic vs. Accidental Gaps

- Yi & Duanmu (2015) proposed four principled Mandarin phonotactic constraints.
  - \*HH: The vowel feature [+high] cannot occur in succession. (e.g., \*[m<sup>h</sup>ui] \*[t<sup>h</sup>yu])
  - \*[Cor]\_[Cor]: [Cor] cannot occur in both on-glide and coda. (e.g., \*[j<sup>h</sup>ai] \*[p<sup>h</sup>jei])
  - \*[Lab]\_[Lab]: [Lab] cannot occur in both on-glide and coda. (e.g., \*[w<sup>h</sup>ou] \*[n<sup>h</sup>wau])
  - Identical articulators cannot occur in succession in onset consonant and on-glide. (e.g., \*[t<sup>h</sup>ʃjan] \*[t<sup>h</sup>ʃye])

# Systematic vs. Accidental Gaps

- All principled constraints are varieties of Obligatory Contour Principle (Leben, 1973; McCarthy, 1986).
- This similarity avoidance effect is grounded in psychological reality because adjacent similar sounds have been shown to present difficulties in production planning (Frisch et al., 2004) and in perception (Woods et al., 2010).
- Typologically, OCP effects in the lexicon are widely reported in many languages such as Arabic (Frisch & Zawaydeh, 2001), Hebrew (Berent & Shimron, 1997), Muna (Coetzee & Pater, 2008), and Quechua (Gallagher, 2010).

# Allophonic Restrictions

- Previous studies mostly discussed phonotactic restrictions held on the **phonemic** level, and few looked into the phonotactic effects of **allophonic** distributions. (e.g. [sp<sup>h</sup>ik] from English ‘speak’, [ʃwe] from Mandarin ‘说’).
- Because allophones of the same phoneme are often categorized under the same category by speakers due to perceptual similarity (Jaeger 1980), and the discrimination of an allophonic distinction is poorer than a phonemic contrast (Whalen et al. 1997).

# Allophonic Restrictions

- But, allophonic distinctions can be reliably heard by speakers.
  - Jia et al. (2006) ask native Mandarin speakers to do ABX discrimination on [æ~ɑ] vowel pair in American English. This contrast is allophonic in Mandarin, yet their discrimination rate reaches 88.9%.
- Therefore, allophonic restrictions will contribute to phonotactic judgment, but it is hypothesized that the violation of such restrictions will not be as fatal.

# Suprasegmental Phonotactics

- In Mandarin, not all existing syllables can freely combine with all of the four tones. Some syllable-tone combinations happen not to exist, e.g. [nei1].
- These missing syllable-tone combinations received significantly lower acceptability than real words in non-word judgment (Kirby & Yu 2007; Myers 2002).
- Co-occurrence patterns on segmental and suprasegmental levels are also likely to be noticed by speakers and form a part of their phonotactic grammar.

# Suprasegmental Phonotactics

- Tones are perceived as conceptually different forms from other segmental level features (Hyman 2011).
- And the processing of lexical tones is disadvantaged compared to segmental information (Cutler & Chen 1997).
- Prediction on acceptability: lower than real words, but higher than other segmental gaps.

# Lexical Statistics: Neighborhood Density

- Speakers may evaluate how similar a word is with other known words.
- One way to quantify this is neighborhood density. It is defined as the number of words generated by substituting, deleting, or adding a single phoneme (Greenberg & Jenkins, 1964).
- Numerous studies have shown that neighborhood density plays an essential role in spoken word perception and production (see Vitevitch & Luce (2016) for a review).
- In acceptability judgment tasks, neighborhood density is positively correlated with ratings (Kirby & Yu, 2007; Myers & Tsay, 2005).

# Research Questions

- Is non-word acceptability judgment gradient in Mandarin?
- How do the grammatical and lexical statistics factors mentioned above contribute to the judgment variation?
- Does grammaticality play an independent role from lexical statistics?
- A syllable well-formedness judgment experiment was conducted to probe into these questions.

# Mandarin Overview

- Mandarin has a (comparably) simple and restrictive syllable structure (namely CGVX), so it is possible to work out a list of all theoretically possible syllables.
  - Onset **C**onsonant: p p<sup>h</sup> m f t t<sup>h</sup> n l ts ts<sup>h</sup> s tʂ tʂ<sup>h</sup> ʂ z tɕ tɕ<sup>h</sup> ɕ k k<sup>h</sup> x
  - **G**lide: j w ɥ
  - Surface **V**owel: i u y e ə o a ɑ
  - **E**Xtra Ending Sound: i u n ŋ
- Mandarin shows rich allophony in vowels.
 

ə → o / w _ #, or _ u	a → a / _ i, n, #
ə → e / j, ɥ _ #, or _ i	a → a / _ u, ŋ
ə → ə / _ n, ŋ, #	a → e / j, ɥ _ n

# Stimuli

- We first listed all theoretically possible Mandarin syllables, both existing and missing syllables.
  - Under the CGVX syllable structure, only the vowel is obligatory. Factorial combination of all sounds gives rise to  $(21+1) * (3+1) * 8 * (4+1) = 3,520$  possible syllables, including 384 existing syllables and 3,136 missing syllables.
  - For English, the number of logically possible syllables is much larger ( $>150,000$ )
- Only the high-level Tone 1 was used.

# Stimuli

- Given that phonotactically illegal forms may cause perceptual difficulties (Dupoux et al. 1999), and this study is only concerned with the acceptability of the syllables that can be reliably identified.
- To that effect, we implemented two steps:
  - 1,273 syllables that could potentially be confusable with other syllables were removed from the list.
  - An AX discrimination pretest was also designed to check whether participants can differentiate the allophonic differences used in the experiment. The overall accuracy was 91.3%.

# Stimuli

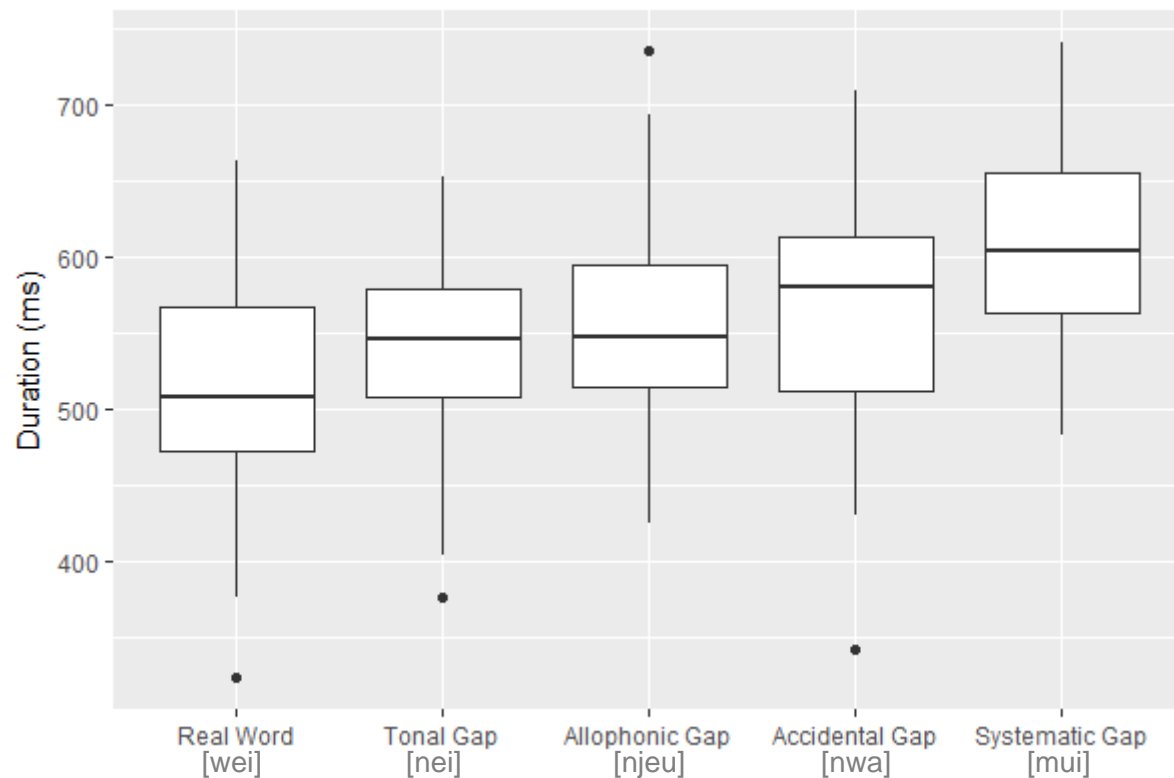
- 384 existing syllables
  - 321 **real words** (e.g., wei)
  - 63 **tonal accidental gaps**, because they cannot take the high-level tone (e.g., nei)
- 1,863 missing syllables (excl. 1,273 perceptually ambiguous forms)
  - 434 **allophonic gaps**, gaps that *only* violate the allophonic rules (e.g., njeu)
  - 1,041 **systematic gaps**, gaps that violate some of the four major phonotactic constraints of Mandarin (e.g., mui)
  - 388 other **segmental accidental gaps**, gaps remain unexplained by the four constraints (e.g., nwa)

# Five Stimulus Types

All Possible Syllables (3,520)					
Existing Syllables (384)		Missing Syllables (3,136)			
Real Words (321) [wei]	Tonal Accidental Gaps (63) [nei]	Allophonic Gaps (434) [njeu]	Segmental Accidental Gaps (388) [nwa]	Systematic Gaps (1,041) [mui]	Forms Indistinguishable from Other Forms (1,273)

- Randomly select 40 syllables from each of the five types as the test stimuli.
- Test stimuli were recorded by a native Beijing Mandarin speaker with phonetic training, and then normalized for intensity.

# Duration by Type



- The stimuli tend to be longer for more ungrammatical stimulus types.

# Participants & Procedure

- 30 native Mandarin speakers born and raised in Northern China. Average age: 24 years old.
- Experiment procedure:
  - 5 practice trials and 200 test trials.
  - Participants heard stimuli from earphones.
  - The instruction asked the participants to rate whether the stimulus sound like Mandarin or not, on a scale from 1 (bad) to 7 (good).
  - Participants clicked on one of the seven rating buttons.
  - 500 ms pause between two trials.
  - Rating responses were recorded.

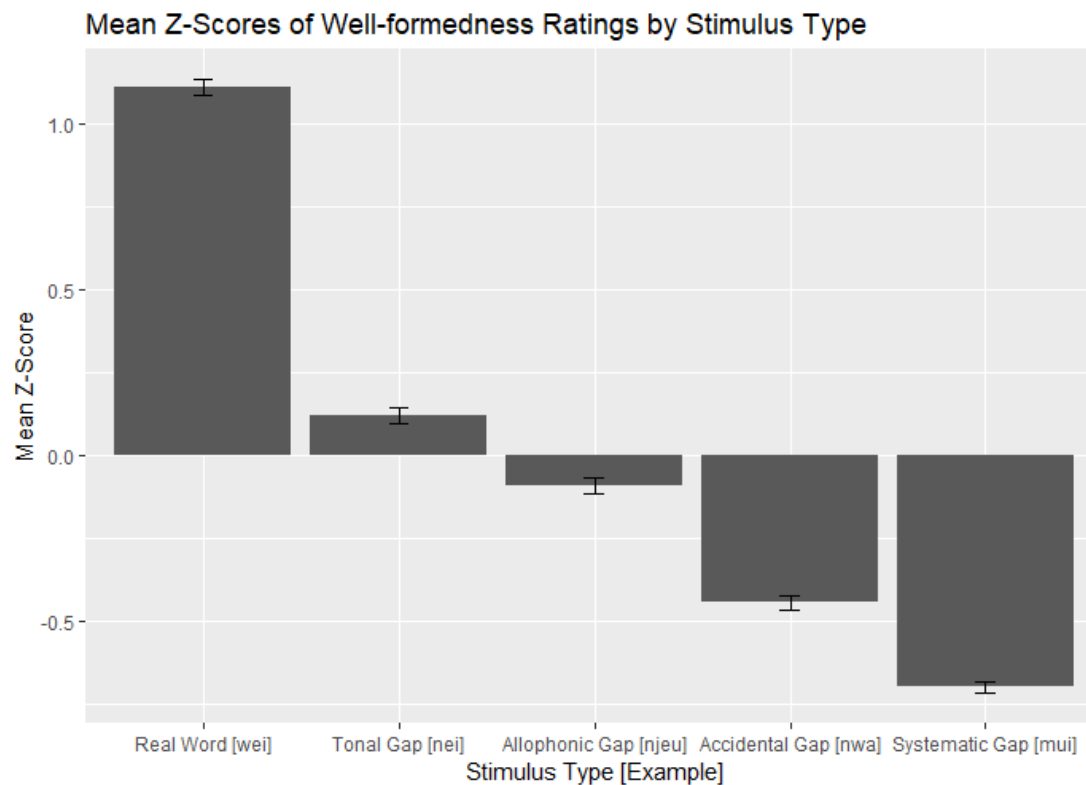
# Data Analysis

- Linear mixed-effects regression model
  - Dependent variable: z-score transformed ratings out of each subject (Casasanto et al. 2010)
  - Random effects: intercept of *Item*
  - Fixed effects: *five stimulus types; neighborhood density; duration*
    - *Duration* was also z-score transformed to avoid excessively distinct scaling

# Data Analysis

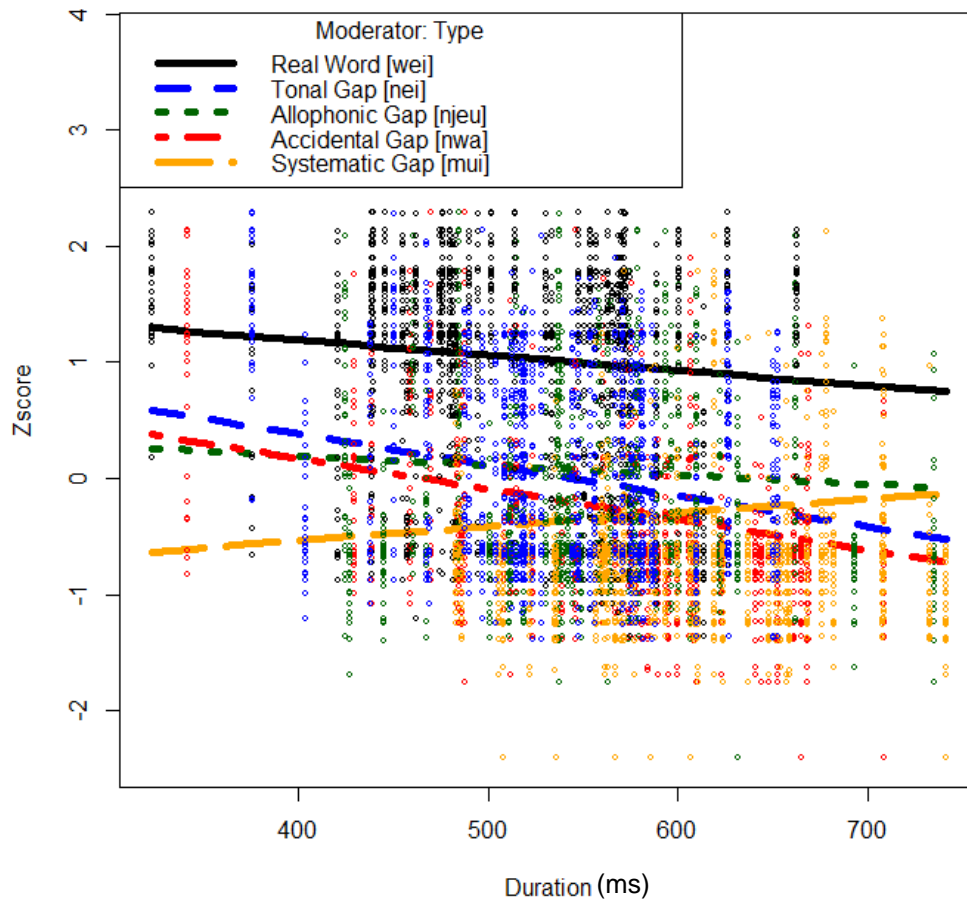
- Both forward and backward step-wise algorithms were attempted for searching for the best model, and the two types of algorithms agreed on the same final model.
- rating  $\sim$  type + ND + duration + type:ND + type:duration + (1|item)

# Rating by Stimulus Type



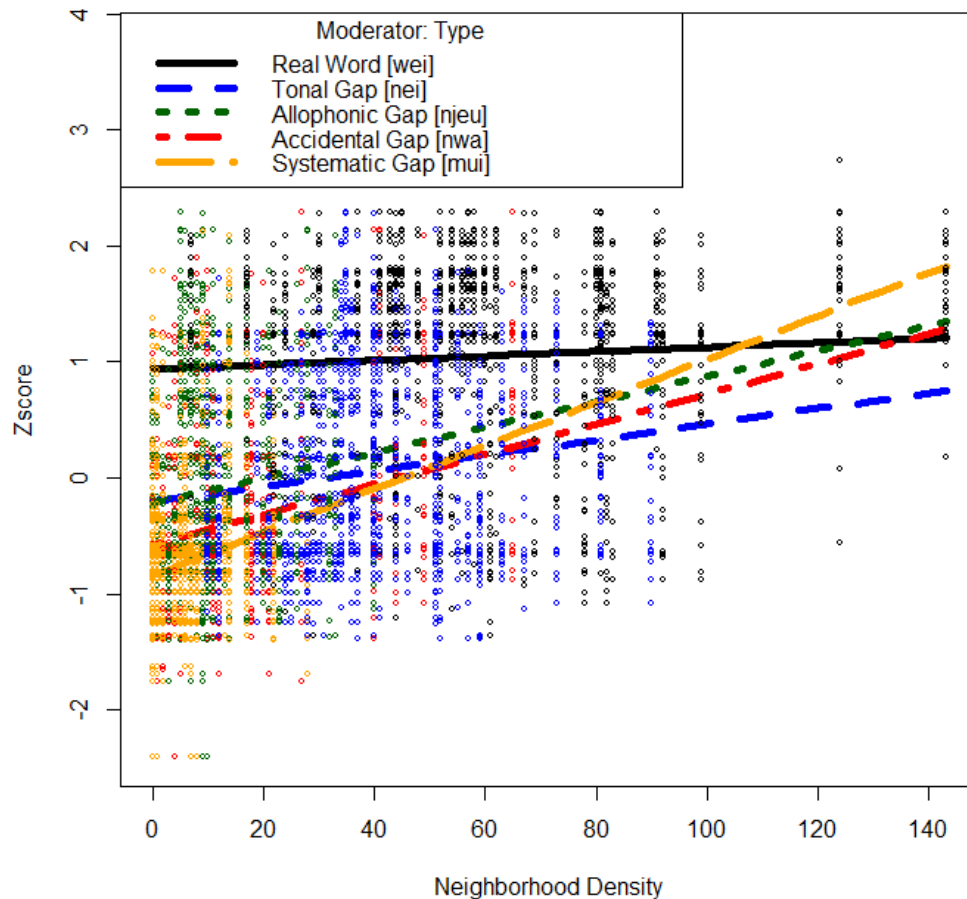
- Post-hoc multiple comparisons with Bonferroni p-value adjustments suggested that the ratings of all five stimulus types were significantly different from each other (all p-values < .0001)

# Stimulus Type : Duration Interaction



- The effect of duration varies for different stimulus types.

# Stimulus Type : ND Interaction



- Neighborhood density, on the other hand, is positively correlated with all five stimulus types overall.
- Less correlation in real words (Kirby & Yu 2007; Myers & Tsay 2005).

# ANOVA Table

	Sum Sq.	Mean Sq.	NumDF	DenDF	F value	p value
Type	51.284	12.8211	4	185	28.5454	<.0001*
Duration	6.905	6.9049	1	185	15.3733	.0001*
Neighborhood Density	4.029	4.0291	1	185	8.9706	.0031*
Type:Duration	4.929	1.2323	4	185	2.7437	.0299*
Type:Neighborhood Density	3.455	0.8636	4	185	1.9228	.1084

- The ANOVA table illustrates that, after duration and neighborhood density are taken into account, the five stimulus types still stand out as the most significant predictor of acceptability judgment.

# Regression without Real Words Data

	Sum Sq.	Mean Sq.	NumDF	DenDF	F value	p value
Type	12.5650	4.1883	3	148	9.6964	<.0001*
Duration	3.2144	3.2144	1	148	7.4416	.0071*
Neighborhood Density	7.0720	7.0720	1	148	16.3723	<.0001*
Type:Duration	5.1745	1.7248	3	148	3.9932	.0091*
Type:Neighborhood Density	1.0594	0.3531	3	148	0.8175	.4861

- This ANOVA table presents the model without real words data.
- The main findings still hold. Stimulus type is still a significant predictor.
- Neighborhood density becomes more significant.

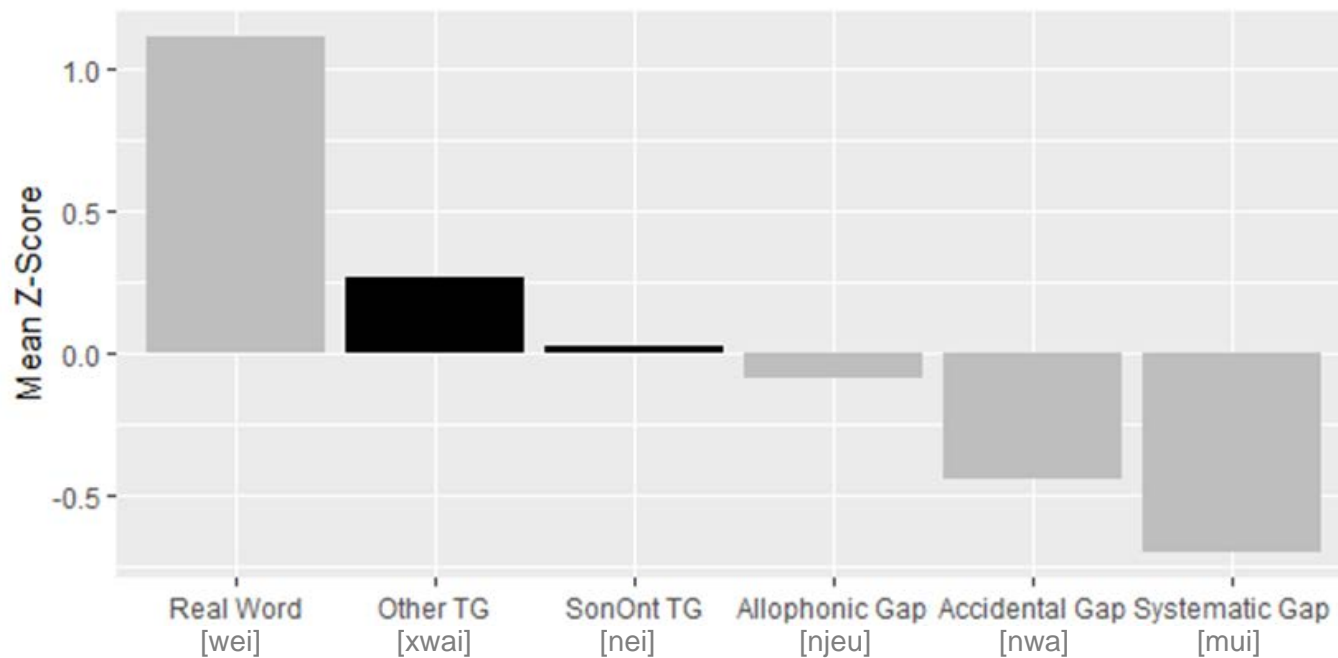
# Further Gradience

- The five stimulus types can be further subcategorized.
- Result in further gradience within each type.

# Tonal Gap

- Among all the 63 Tone 1 tonal gaps, 40 of them start with a sonorant onset [m n l z].
- This is because sonorant onset syllables with Tone 1 lack historical sources.
  - Modern Mandarin sonorant onsets [m n l z] are predominantly derived from Mid Chinese voiced sounds. Yet, the syllables carrying Tone 1 in modern Mandarin develop from Mid Chinese syllables with voiceless onsets only.
- Rating results suggest that speakers are aware of this trend in the lexicon.

# Tonal Gap

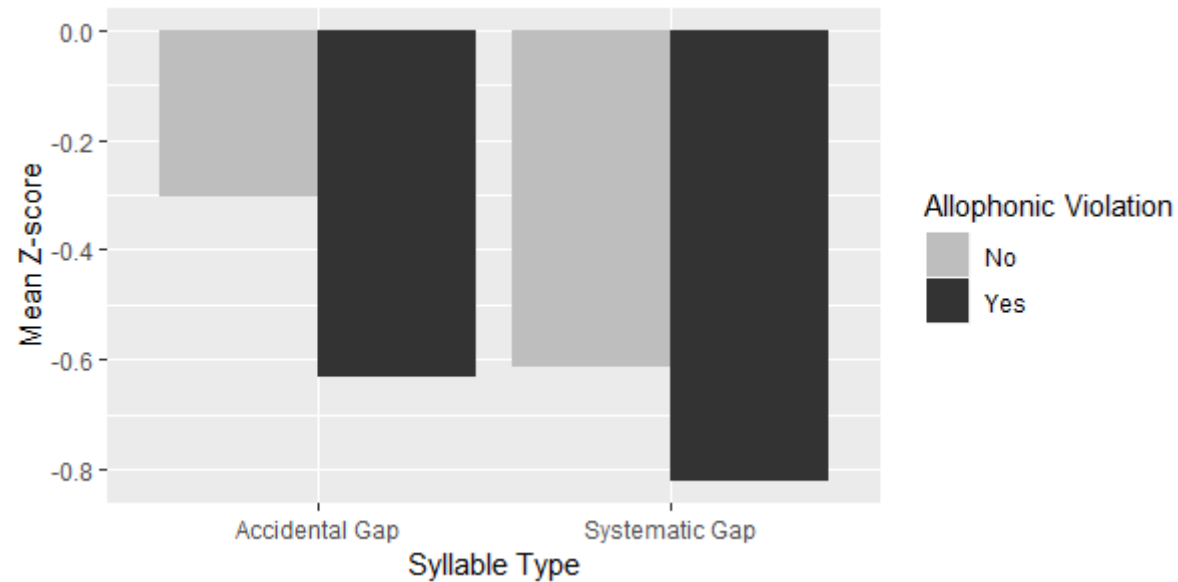


Two-tailed independent-samples t-test suggested that the difference is significant ( $t(1198) = -4.9588, p < .0001$ ).

# Additive Effect

- Systematic gaps and accidental gaps may or may not violate allophonic restrictions.
  - For example, the systematic gap [nwau] not only violates the labial cooccurrence constraint, but also is allophonically inappropriate since the low vowel /a/ before the off-glide [u] should surface as the back [ɑ], not the front [a].
- Rating data suggest that gaps that also violate allophonic restrictions are judged worse than those who do not.

# Additive Effect



# Summary

- Mandarin non-word judgment is gradient: Real words > Tonal Gaps > Allophonic Gaps > Accidental Gaps > Systematic Gaps
- Multiple grammatical factors contribute to the gradient, including OCP-based phonotactic constraints, allophonic restrictions, syllable-tone combination restrictions.
- Lexical statistics contributes to the gradient acceptability as well. But, grammatical factors play an independent role and explain a large amount of variation in judgment.

Thank you!

# References

- Bailey, T. M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory & Language*, 44(4), 568.
- Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition*, 64, 39–72.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3), 591–630.
- Chomsky, N., & Halle, M. (1965). Some controversial questions in phonetical theory. *Journal of Linguistics*, 1(2), 97–138.
- Coetzee, A. W., & Pater, J. (2008). Weighted Constraints and Gradient Restrictions on Place Co-Occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, 26(2), 289–337.
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59(2), 165–179.
- Duanmu, S., & Yi, L. (2015). Phonemes, Features, and Syllables: Converting Onset and Rime Inventories to Consonants and Vowels. *Language and Linguistics*, 16(6), 819–842.
- Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*.
- Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22(1), 179–228.
- Frisch, S. A., & Zawaydeh, B. A. (2001). The Psychological Reality of OCP - Place in Arabic. *Language*, 77(1), 91–106.
- Gallagher, G. (2010). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology*, 27(3), 435–480.
- Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, 20(July), 157–177.

# References

- Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, 44(1), 45–75.
- Hyman, L. M. (2011). Tone: Is it different? In J. A. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory: Second Edition* (pp. 197–239). Oxford: Blackwell.
- Jaeger, J. J. (1980). Testing the Psychological Reality of Phonemes. *Language and Speech*, 23(3), 233–253.
- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119(2), 1118.
- Kirby, J. P., & Yu, A. C. L. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 21(August), 1389–1392.
- Leben, W. (1973). *Suprasegmental Phonology*. MIT.
- McCarthy, J. J. (1986). OCP Effects: Gemination and Antigemination. *Linguistic Inquiry*, 17(2), 207–263.
- Myers, J. (2002). An analogical approach to the Mandarin syllabary. *Journal of Chinese Phonology*, 11(Special Issue), 163–190.
- Myers, J., & Tsay, J. (2005). The processing of phonological acceptability judgments. In *Proceedings of symposium on 90-92 NSC projects* (pp. 26–45).
- Shademan, S. (2007). Grammar and analogy in phonotactic well-formedness judgments. *Dissertation*, 174.
- Vitevitch, M. S., & Luce, P. A. (2016). Phonological Neighborhood Effects in Spoken Word Perception and Production. *Annual Review of Linguistics*, 2(1), annurev-linguist-030514-124832.
- Woods, D. L., Yund, E. W., Herron, T. J., & Cruadhlaioich, M. A. I. U. (2010). Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *The Journal of the Acoustical Society of America*, 127(3), 1609–1623.

## More on Articulator Dissimilation

- Dentals [t t<sup>h</sup> n l ts ts<sup>h</sup> s] are allowed to combine with the coronal glides [j ɥ], e.g., [t<sup>h</sup>jan] ‘sky’, because [Coronal] is underspecified in dentals and they receive the [Coronal] feature redundantly.

# Stimuli

- However, not all syllables will be considered. Many sequences will very likely be misperceived as other existing forms or hard to distinguish from each other. I propose following criteria to rule out ambiguous forms.
  - No glide distinction before [y]: all glides before the vowel [y] are considered neutralized, i.e., [jy]=[wy]=[ɥy]. Only [jy] was preserved in the possible syllable list.
  - No [+round] distinction before [u]: the glides [j] and [ɥ] before the vowel [u] are considered neutralized, i.e., [ju]=[ɥu]. Only [ju] was preserved in the possible syllable list.
  - No distinction between [tɕ] and [tɕj] or between [tɕw] and [tɕɥ]; only [tɕ] and [tɕw] were preserved in the possible syllable list.
  - No distinction between [oŋ] and [uŋ]. Only [uŋ] was preserved in the possible syllable list.
  - No distinction between [an] and [aŋ], or between [aŋ] and [aŋ]. Only [an] and [aŋ] were preserved in the possible syllable list.
- These criteria mark 1,273 syllables as undistinguishable from some other syllables. Therefore, the remaining list contains 1,863 missing syllables and 384 existing syllables.

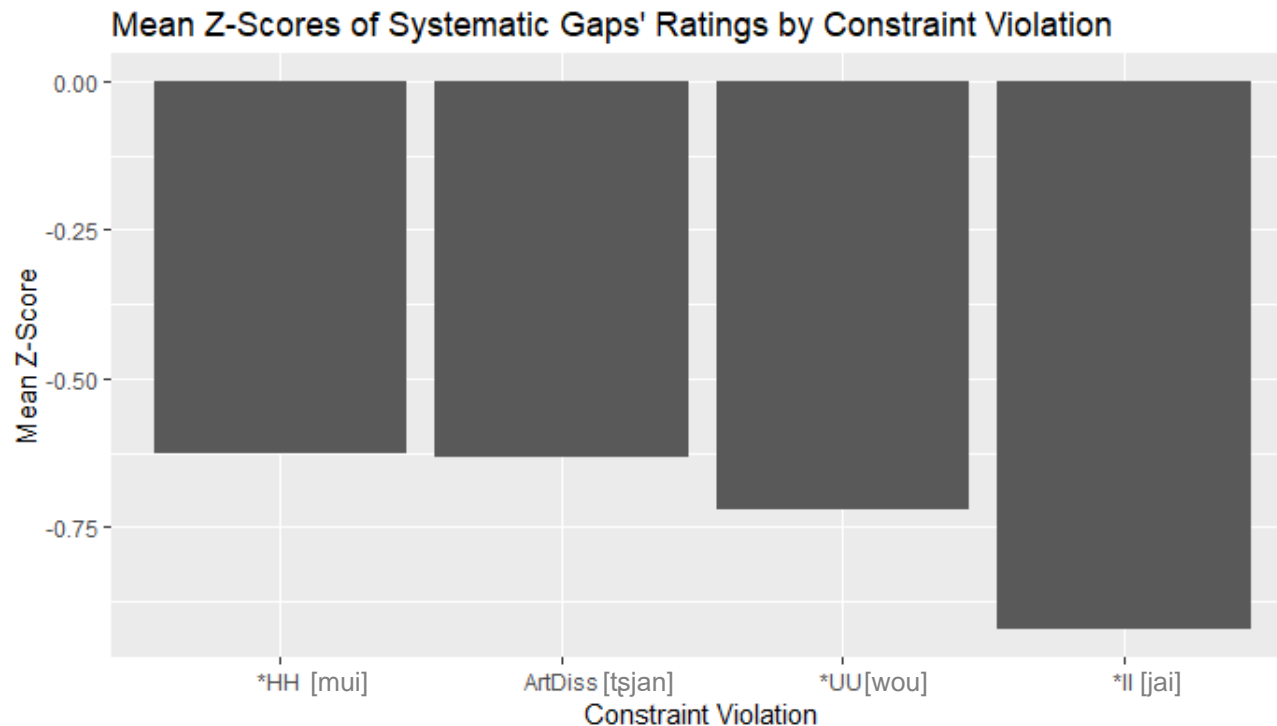
# Results

	Estimate	Std Error	t value	p value
(Intercept)	0.9396	0.1449	6.485	<.0001*
<i>Tonal Gap</i>	-1.1405	0.2023	-5.637	<.0001*
<i>Allophonic Gap</i>	-1.1582	0.1720	-6.734	<.0001*
<i>Accidental Gap</i>	-1.5142	0.1697	-8.924	<.0001*
<i>Systematic Gap</i>	-1.7807	0.1746	-10.198	<.0001*
<i>Duration</i>	-0.0982	0.0675	-1.455	.1472
<i>Neighborhood Density</i>	0.0019	0.0023	0.826	.4100
<i>Tonal Gap : Duration</i>	-0.1089	0.0993	-1.096	.2744
<i>Allophonic Gap : Duration</i>	0.0371	0.0956	0.388	.6988
<i>Accidental Gap : Duration</i>	-0.0983	0.0983	-1.000	.3186
<i>Systematic Gap : Duration</i>	0.1881	0.0998	1.886	.0609
<i>Tonal Gap : ND</i>	0.0048	0.0039	1.256	.2106
<i>Allophonic Gap : ND</i>	0.0091	0.0066	1.367	.1733
<i>Accidental Gap : ND</i>	0.0111	0.0053	2.089	.0381*
<i>Systematic Gap : ND</i>	0.0168	0.0010	1.682	.0942

# Systematic Gap

- Plot the mean acceptability of systematic gaps by their constraint violation.
  - Only violate \*HH (e.g., [mui])
  - Only violate \*[Cor]\_[Cor] (e.g., [jai])
  - Only violate \*[Lab]\_[Lab] (e.g., [wou])
  - Only violate Articulator Dissimilation (e.g., [tʃjan])

# Systematic Gap



Pairwise *t*-tests with Bonferroni *p*-value adjustments suggested that the ratings of the gaps violating \*ll were significantly different from the gaps violating \*HH ( $p < .0001$ ) and from the gaps violating ArtDiss ( $p = .0006$ ), and the rating difference between \*ll and \*UU was marginally significant ( $p = .0512$ ). All other pairs did not reach statistical significance.

## (Absence of) Additive Effect

- Systematic gaps can either violate only one of the principled phonotactic constraints or violate more than one constraint.
- Multiple constraint violations do not necessarily lead to lower acceptability ratings (Kirby & Yu 2007).

# (Absence of) Additive Effect

