Interspeech 2024
1-5 September 2024, Kos, Greece

# Perceptual Learning in Lexical Tone: Phonetic Similarity vs. Phonological Categories

*Ariëlle Reitsema*[1], *Chenxin Li*[1], *Leanne van Lambalgen*[2], *Laura Preining*[1], *Saskia Galindo Jong*[2], *Qing Yang*[1], *Xinyi Wen*[1], *Yiya Chen*[1]

[1]Leiden University Centre for Linguistics, Leiden University, the Netherlands
[2]University of Amsterdam, the Netherlands
a.reitsema@umail.leidenuniv.nl, chenxin.li@hotmail.com, ljvl1210@hotmail.com, laura.preining@hotmail.com, saskiagalindo@yahoo.com, q.yang@hum.leidenuniv.nl, x.wen@hum.leidenuniv.nl, yiya.chen@hum.leidenuniv.nl

## Abstract

In speech comprehension, listeners recalibrate their interpretation of variable speech signals through exposure and disambiguating information. Recalibration is attested both segmentally and suprasegmentally, but little is known about what constrains it in lexical tone. This project investigated the effects of phonological categories and phonetic similarity on perceptual learning. We exposed Chinese listeners to pitch contours ambiguous between two tone categories (realised with a level or rising pitch contour) and lexically biased their perception to one interpretation. Crucially, the rising pitch contour could be from two different phonological tone categories. Perceptual learning was observed not only in the rising contours used for exposure but also across phonetically similar but phonologically different rising contours, suggesting that perceptual learning in tone is not constrained by phonological tone categories but is facilitated by the phonetic similarity of pitch contours.

**Index Terms**: speech perception, lexical tone, perceptual learning, Standard Chinese, phonological categories, phonetic similarity

## 1. Introduction

Listeners rapidly and successfully adjust their perception of variable speech signals. They can use supplementary information, such as knowledge of the lexicon [1], lip movements [2], or other contextual cues [3] to infer how they should categorise an ambiguous speech sound, and can generalise this learning to new contexts through sufficient exposure. For example, when repeatedly confronted with s-final words (e.g., *glass*) where the /s/ is replaced by a sound ambiguous between [s] and [f], listeners learn to associate this ambiguous sound with /s/ due to lexical bias, since /glaf/ is not a word. Their perception of such ambiguous sounds is subsequently skewed towards /s/ [1], also in lexically ambiguous contexts (e.g., *lice* – *life*) that they have not previously encountered [4]. This phenomenon is known as lexically guided perceptual recalibration.

Given that recalibration effects generalise to new contexts, it has been argued that abstraction plays a role in perceptual learning [4], [5], [6]. Recalibration can then be seen as an adjustment in the mapping between the auditory speech signal and abstract speech sound representations. However, there is a discussion in the literature on exactly what form of abstraction guides and constrains the generalisation of recalibration. A few experimental studies have attempted to gather support for different linguistic units that could function in perceptual learning, such as phonemes [7], phonological features [8], and allophones [9], yet there is no consensus on one particular linguistic unit of prelexical representation.

Instead, support for a panoply of different units that could underlie perceptual learning has been generated [10], [11], [12]. Still, counter-evidence due to conflicting data demonstrates the shortcomings of these linguistic categories in outlining the extent of perceptual learning. Crucially, experimental data suggest that lower-level information, e.g., acoustic similarity, can at times trump higher-level linguistic categories in perceptual learning [10]. For instance, Bowers, Kazanina, and Andermane [7] made strong claims in favour of phonemes based on findings of selective adaptation effects across the same phoneme in different syllable positions. However, Samuel [10] did not find the same effects in a replication where acoustic similarity due to plosive release noise had been reduced.

Similarly, Kraljic and Samuel [8] found that exposure to ambiguous /t/ or /d/ phonemes in lexically biasing contexts did not only recalibrate the /t/ – /d/ contrast, but also affected the /p/ – /b/ contrast, which has a different place of articulation but shares the same [± voice] feature. While they conclude that recalibration takes place at the level of phonological features, Mitterer et al. suggest that the observed cross-categorical generalisation was rather due to phonetic similarity (namely, post-plosive aspiration noise) between the realisations [3, p. 112]. They further support the role of phonetic similarity with their study on recalibration in /plain/ and /tense/ stops in Korean, where /plain/ stops are realised as [tense] in some phonetic contexts. They observed generalisation to underlyingly different representations (/tense/ to /plain/ and vice versa), but only if these were phonetically similar ([tense]). They also observed generalisation to phonetically different realisations ([tense] to [plain] and vice versa), but only if these had the same phonemic underlying representation (/plain/), although this effect was less robust.

The role of the underlying phoneme in perceptual learning was further questioned by Mitterer et al. [9]. They observed that recalibration of one allophonic variant of the Dutch /r/ – /l/ contrast (e.g., [r] – [l]) did not affect other allophonic implementations of this same phonemic contrast (e.g., [ɹ] – [ɬ]). Although they propose that context-specific allophones constrain recalibration, they also harken back to the relevance of phonetic similarity by stating that "listeners' generalisations


10.21437/Interspeech.2024-2212

of perceptual learning are tightly bound to the acoustic patterns they experience" [9, p. 360]. There is even evidence supporting context-specific recalibration, where generalisation only occurs in identical phonetic contexts. Reinisch et al. [13] examined learning on the /b/ – /d/ contrast and found that learning did not generalise from [aba] – [ada] to [ibi] – [idi].

Overall, there is little consensus regarding a single linguistic unit playing the lead role in constraining and facilitating recalibration. Moreover, existing literature suggests that both higher-level linguistic categories, lower-level acoustic information, and the linguistic context, play significant roles in governing perceptual retuning in speech processing.

The current study applies the lexically guided perceptual learning paradigm to lexical tone, to investigate the role of phonetic similarity and phonological categories in facilitating or constraining the generalisation of tone recalibration. Previous perceptual learning studies focused on the segmental domain, but recalibration was also observed suprasegmentally, for example in intonation contours [14], lexical stress [15], and lexical tones . Although lexical tone contrasts are cued differently from segmental contrasts (their acoustic cues usually extend over a full syllable [16]), Mitterer, Chen and Zhou [6] found recalibration effects in Standard Chinese tones to be analogous to those observed segmentally. They exposed listeners to tone contours manipulated to be ambiguous between tone 1 (level; T1) and tone 2 (rising; T2) in lexically biased contexts. In the test phase, they observed recalibration effects in both previously presented and newly encountered monosyllabic items. The effects were slightly stronger for the familiar items, but generalisation to new words indicates that abstraction concerning lexical tone category information plays a role. Yet, little is known about further factors that facilitate and constrain the generalisation of lexical tone recalibration.

Building on Mitterer et al. [6], the current study replicated their observation of lexical tone recalibration and extended the investigation of its spread to underlyingly different but phonetically similar tones in a bisyllabic tonal context. To this end, we made use of tone 3 sandhi, a phonological process in Standard Mandarin Chinese (Putonghua) where tone 3 (low; T3) followed by another T3 is realised with a rising f0 contour, which is comparable to that of T2 (see 1-2) [17]. Although sandhi T3 and T2 are not phonetically identical [17], they are similar enough that native speakers typically fail to distinguish them in perception [18]. (See, e.g., [19], [20] for further details on the representation and processing of T3 sandhi.)

(1) T3 sandhi:
T3 → Rising tone /__T3

(2) /jing3/ + /dian3/ → [jing3$^{rising}$ dian3]

We manipulated tone contours to be ambiguous between T1 and T2 (transcribed hereafter as [½]), as well as T1 and sandhi T3 ([⅓]). Using an exposure-test design, we presented syllables with these manipulated contours in lexically biasing contexts in the exposure phase and in lexically ambiguous contexts in the test phase. To induce recalibration during the exposure phase, one tone contour in a contrast was consistently replaced by a manipulated contour, while the other was left unaltered. For example, to bias a participant's perception of the T1–T2 contrast towards T2, we consistently replaced T2 contours with [½] in contexts that lexically favoured a T2 interpretation, while presenting all instances of T1 naturally. By testing whether perceptual learning in T2 generalises to sandhi T3 and vice versa, we can examine whether generalisation only occurs within phonological categories or can also occur over phonetically similar but phonologically different tones.

# 2. Methods

## 2.1. Participants

140 native speakers of Standard Chinese, aged between 18 and 50 (M = 23.4), completed the online experiment for compensation. They were recruited via Chinese social media platforms (*Weibo*; *WeChat*), at Leiden University, and through personal networks. All participants grew up in northeastern regions of China, where the local dialects, like Standard Chinese, belong to the Mandarin family.[1] According to their self-reports, about half of the participants spoke another local Mandarin dialect, but all had acquired Standard Chinese as a young child, at the latest since primary school.

## 2.2. Materials

The experimental materials consisted of auditory stimuli accompanied by visual stimuli. The auditory stimuli recordings were made by a male native speaker of Standard Chinese who was born and raised in Beijing. For the recordings, we used the sentence frame given in (3), where the empty slot was filled with bisyllabic targets. These target words were selected from the SUBTLEX-CH database, taking into account word frequency.[2] The visual stimuli were target words clearly displayed on screen in Chinese characters.

(3) *ta1 shuo1 __ zhe4 ge4 ci2*
'(s)he said the word __'

For the first syllable of the exposure phase targets, we selected 40 T1–T2 minimal pairs and 40 T1–T3 minimal pairs. The second syllable of each target was T3 so that any T3 contours in the first syllable would be realised with a rising contour. This second syllable differed segmentally between two members of a pair, serving to disambiguate which lexical tone was intended on the first syllable. Sample target pairs for the two exposure conditions are shown in (4) and (5). An additional 40 pairs of filler items were recorded in the same sentence frame and were used in both exposure conditions. To resemble the structure of the target items, the first syllable of each pair was an identical tone 4 (falling; T4) syllable, while the second syllable differed both segmentally and suprasegmentally, alternating between T1 and T2 (6).

Sample target pairs for the exposure phase:

(4) T1 – T2 exposure pairs (n = 40):
*tian1 shi3* 'angel' – *tian2 pin3* 'dessert'

(5) T1 – T3 exposure pairs (n = 40):
*yin1 ying3* 'shadow' – *yin3 pin3* 'beverage'

(6) Fillers (n = 40):
*da4 jia1* 'everyone' – *da4 xue2* 'university'

The test phase targets were bisyllabic minimal pairs, crucially differing only in the lexical tone of the first syllable (see 7 – 8). The second syllable was identical for both members of each pair

[1] Participants grew up in the following regions: Beijing, Tianjin, Hebei Province, Henan Province, Shandong Province, Heilongjiang Province, Jilin Province, or Liaoning Province.

[2] There is no consensus on what constitutes a word in Chinese; we selected bisyllabic items that are frequent and commonly considered lexical collocations, if not words.

to prevent lexical disambiguation. All carried T3. We selected 20 T1–T2 minimal pairs and 20 T1–T3 minimal pairs, none of which had been used in the exposure phase.

Sample target minimal pairs for the test phase:

(7) T1 – T2 test pairs (n = 20)
*chu1xue3* 'first snow' – *chu2xue3* 'snow removal'

(8) T1 – T3 test pairs (n = 20)
*qi1 wu3* 'to bully' – *qi3 wu3* 'to dance'

Ambiguous tone contours for the exposure and test phase targets were created using *Praat* [21]. The first syllables of both members of a pair were manually selected and their duration was normalised (i.e., set to the average of the two syllables). F0 contours were estimated via pitch points, which were interpolated to create five steps on T1–T2 and T1–T3 continua, as illustrated in Figure 1. Finally, the original recordings were re-synthesised by replacing the original f0 contours with the interpolations using the Pitch-Synchronous-Overlap-and-Add (PSOLA) method. The midpoint ambiguous contour (50% T1; 50% T2 or sandhi T3) was used in the exposure phase. In the test phase, the three intermediate steps (25%, 50%, 75% T1) from the f0 continua replaced the original f0 contours on the T1 recordings. The unambiguous endpoints of the continua were excluded since the goal was to test the perception of ambiguous tone contours (see also [1], [6]).

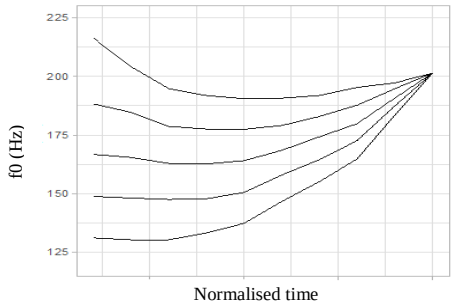


Figure 1: *5-step f0 continuum between the pitch curves of 'qi1' and 'qi3'*

### 2.3. Apparatus and procedure

The experiment was created in the Gorilla Experiment Builder (www.gorilla.sc) so that the experiment could be hosted remotely [22]. Participants were instructed to respond to trials by clicking with the mouse. After passing an equipment check and receiving written instructions, participants were randomly assigned to one of four exposure conditions listed in Table 1. Half of the participants heard [½], while the other half heard [⅓]. Within the first group, one subgroup was biased to interpret the ambiguous contour as T1, while the other subgroup was biased to interpret the ambiguous contour as T2. Analogously, the second group also had a T1 and T3 bias group, respectively.

In the exposure phase trials, an auditory stimulus was played immediately following the onset of an accompanying visual target displayed as two Chinese characters. Participants were instructed to click on the on-screen button if they heard the visually displayed characters in the audio. Participants had 5000ms (from the onset of the trial) to respond, while a time-out was a negative response. Each participant reacted to a total of 160 randomly ordered trials, composed of 40 T1 items and 40 T2 or T3 items, depending on the condition the participant was assigned to, as well as 80 fillers. The visual and auditory targets matched in approximately half of the trials of each type.

Table 1: *Between-subjects bias conditions in the exposure phase. T3 or [3] refers to sandhi T3.*

| Exposure contrast | Bias condition group | /level contour/ | /rising contour/ |
|---|---|---|---|
| T1 – T2 | level | [½] | [2] |
| | rising | [1] | [½] |
| T1 – T3 | level | [⅓] | [3] |
| | sandhi rising | [1] | [⅓] |

The test phase was set up as a forced-choice identification task. Both members of bisyllabic T1T3–T2T3 and T1T3–T3T3 minimal pairs were present on screen simultaneously with an auditory stimulus in which the target was ambiguous on the T1–T2 or T1–T3 continuum. Participants had to indicate with a mouse-click which one they perceived. There were 120 trials, consisting of 60 T1T3–T2T3 trials and 60 T1T3–T3T3 trials (20 minimal pairs × 3 continuum steps each), none of which had occurred in the exposure phase. These were presented in a random order to each participant and the location of minimal pair members on the screen (left or right) was also randomised.

### 2.4. Analysis

Categorisation responses were analysed using the generalised linear mixed-effects model (GLMM) with the binomial family. All the statistical analyses were run in *R Studio* [23] with the *lme4* package [24]. The binomial dependent variable was participants' categorisation of the test stimuli as having either a rising f0 contour (for either T2 or sandhi T3) or a level f0 contour (for T1) in the first syllable. A maximum model was constructed first, including fixed effects of Test Pair (T1T3–T2T3 vs. T1T3–T3T3), Bias Group (level-bias and rising-bias with T1T3–T2T3 and T1T3–T3T3 exposure), Step (Step2, 3, 4 of the stimulus f0 contours) and the interaction between all fixed factors, as well as by-subject and by-item random intercepts and random slopes for each fixed term. We employed the package *buildmer* [25] to find the maximum model that converged and performed stepwise elimination on each factor.

## 3. Results

As shown in Figure 2, categorisation data in the test phase were visualised by calculating the proportions of the rising tone responses. As we can see, higher steps on the f0 continuum (i.e., manipulated more towards the rising pitch contour) led to more rising tone responses, regardless of whether the responses were T2 in T2T3 or sandhi T3 in T3T3. The left two panels present the proportion of rising tone responses in categorising T1T3–T2T3 minimal pairs, while the right two panels illustrate the results in categorising T1T3 versus T3T3. The upper two panels – which appear visually identical, but do not plot individual variation – present the proportion of rising tone responses when participants were biased to interpret the ambiguous contour (in the exposure) as either T2 (i.e., rising-bias; dotted line) or T1 (i.e., level-bias; solid line). The lower two panels show results of participants who were biased to interpret the ambiguous exposure contour as either T3 (i.e., sandhi rising-bias; dotted line) or T1 (i.e., level-bias; solid line). In each of the four panels, there is a clear difference between the level-bias and the rising-bias groups. As the test stimuli were more likely to be perceived as a rising contour (for lexical T2 or sandhi T3) by the rising-bias groups, these patterns indicate the effects of recalibration across different exposure groups and test pairs.

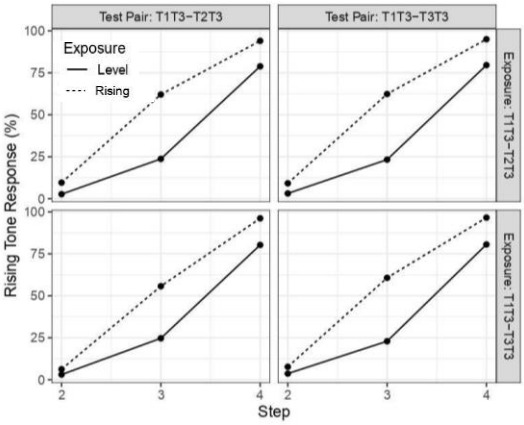


Figure 2: *Proportion of rising tone responses in the test phase.*

The final GLMM consisted of fixed effects of Bias Group, Step, the interaction between Bias Group and Step, and the random intercept for Subject. Test Pair and its interaction with Bias Group and Step did not significantly improve model fit ($p = 0.212$). As there was a significant interaction between Bias Group and Step ($p < 0.001$), pairwise comparisons between Bias Group and Step were computed using the R package *emmeans* [26]. The holm-Bonferroni method was implemented to correct family-wise errors [27]. According to the model summary, there were significantly more rising responses in the rising-bias groups than in the level-bias groups at each step, regardless of whether they were exposed to a rising f0 with T2T3 or T3T3 (all $p < 0.001$). This shows clear evidence for recalibration effects. Moreover, there was no significant difference between responses of the rising-bias in T2T3 and that in T3T3 at each step (Step2, $p = 0.135$; Step3, $p = 1$; level-bias: Step4, $p = 1$). Overall, we observed the recalibration of pitch cues to lexical tone in Standard Chinese. Importantly, the same categorisation pattern between T1T3–T2T3 and T1T3–T3T3 exposure groups, as well as the absence of a Test Pair effect, indicated that the recalibration of tonal cues is not significantly affected by the lexical tonal categories.

## 4. Discussion and conclusion

This study investigated the role of phonological categories and phonetic similarity in perceptual learning in Standard Chinese tones. We found that exposure to an ambiguous tone contour in a lexically biasing context led to perceptual learning on that tonal contrast. Participants biased to perceive pitch contours ambiguous between T1 and T2 ([½]) as a rising contour (for lexical T2) also generalised this learning to contours ambiguous between T1 and sandhi T3 ([⅓]) and vice versa. Thus, perceptual learning was not only generalised to new syllables with the same tonal contrast, but also to phonetically similar contours belonging to phonologically different tonal categories.

These findings serve as a replication of lexical tone recalibration, as observed by Mitterer et al. [6], since the recalibration effect was generalised from exposure phase to test phase syllables. This supports the idea that listeners can equally adapt to segmental and suprasegmental variability and make use of abstraction [15], [28]. Furthermore, our results extended the findings to a bisyllabic context. We replicate the perceptual recalibration effect of tone in the word-initial position, where perceptual recalibration was found to be inhibited, at least in the segmental domain [28]. This might be because tone contours extend across a full syllable, thus differing from the limited temporal nature of a word-initial segment.

Our results are informative about the role of phonological categories in recalibration. We had expected that recalibration effects would be restricted to the phonological contrast participants heard in the exposure phase if phonological categories constrain perceptual learning. However, in our findings, recalibration of one rising contour spread cross-categorically to the acoustically similar but phonologically different tones (i.e., lexical T2 rising and T3 sandhi rising) that were not previously encountered. This suggests that recalibration is not constrained by phonological categories but is instead facilitated by acoustic similarity.

This echoes findings from segmental studies, e.g., Mitterer et al. [3], who similarly observed cross-categorical recalibration of stops in Korean, given close phonetic resemblance. In addition, they found that recalibration effects could spread to phonetically different allophones of the same phoneme, though with a weaker effect. Given that the T3 sandhi may be considered allophonic with T2, a follow-up experiment could investigate further whether phonetic similarity in lexical tones does not only play a facilitating role but also a constraining role (such that recalibration does not spread to categorically the same but phonetically dissimilar tones).

Observations from perceptual learning experiments can shed further light on what forms of abstraction are used in pre-lexical processing. While there have been efforts to find out if linguistic units such as phonemes and allophones play a role in pre-lexical processing, the idea that one particular linguistic unit plays such a role in the first place may be ungrounded [10]. Our findings are consistent with the view that recalibration makes reference to phonetically definable units. Mitterer et al. [3] propose that these can be "fine-grained phonetic events" such as "aspiration noise" or "short noise burst." In lexical tone, generalisation then operates on phonetically definable properties such as f0 contours. This resonates with Scott's [29], findings that recalibration effects can be ear-specific if induced by presenting ambiguous sounds to one ear and unambiguous sounds to the other. This is interpreted as an indication that recalibration "targets low-level, possibly pre-linguistic, sound representations" [29, p. 165]. Our observation that lower-level information seems to play a role also makes sense in light of Samuel's view that higher-level "linguistic units, such as phonemes or allophones, do not have any privileged status in the process of spoken word recognition" [10, p. 46].

Overall, the current research provides strong evidence for the role of phonetic similarity, rather than phonological categories, in facilitating perceptual learning. More research is necessary to clarify the criteria for defining phonetic similarity and to determine to what extent it limits recalibration. Our results do not rule out the facilitating role of more abstract representations; they raise questions about the sufficiency and necessity of conditions under which phonetic similarity and abstract phonological representations facilitate or constrain perceptual recalibration. Furthermore, a cautionary note is warranted, as listeners have been reported to respond differently to artificial and natural stimuli [30]. Therefore, the recalibration effects induced by the carefully tailored stimuli in the current study may not reflect fully how listeners adapt to variable speech in natural contexts. Future research is needed to gain a comprehensive understanding of how listeners cope with variability in speech.

## 5. Acknowledgements

We would like to express our gratitude to Noa de Lange for being an integral part of this research project, and to Jos Pacilly for his assistance and support in developing the code to manipulate the stimuli in *Praat*. We would also like to thankfully acknowledge the financial support by the Leiden University Centre for Linguistics and the Netherlands Organization for Scientific Research (Vici Grant No. vi.c.181.040 to Yiya Chen).

## 6. References

[1] D. Norris, J. M. McQueen, and A. Cutler, 'Perceptual learning in speech', *Cognitive psychology*, vol. 47, no. 2, pp. 204–238, 2003, doi: 10.1016/S0010-0285(03)00006-9.

[2] P. Bertelson, J. Vroomen, and B. de Gelder, 'Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect', *Psychol Sci*, vol. 14, no. 6, pp. 592–597, Nov. 2003, doi: 10.1046/j.0956-7976.2003.psci_1470.x.

[3] H. Mitterer, T. Cho, and S. Kim, 'What are the letters of speech? Testing the role of phonological specification and phonetic similarity in perceptual learning', *Journal of Phonetics*, vol. 56, pp. 110–123, May 2016, doi: 10.1016/j.wocn.2016.03.001.

[4] J. M. McQueen, A. Cutler, and D. Norris, 'Phonological Abstraction in the Mental Lexicon', *Cognitive Science*, vol. 30, no. 6, pp. 1113–1126, 2006, doi: 10.1207/s15516709cog0000_79.

[5] H. Mitterer and E. Reinisch, 'Surface forms trump underlying representations in functional generalisations in speech perception: the case of German devoiced stops', *Language, cognition and neuroscience*, vol. 32, no. 9, pp. 1133–1147, 2017, doi: 10.1080/23273798.2017.1286361.

[6] H. Mitterer, Y. Chen, and X. Zhou, 'Phonological Abstraction in Processing Lexical-Tone Variation: Evidence From a Learning Paradigm', *Cognitive Science*, vol. 35, no. 1, pp. 184–197, 2011, doi: 10.1111/j.1551-6709.2010.01140.x.

[7] J. S. Bowers, N. Kazanina, and N. Andermane, 'Spoken word identification involves accessing position invariant phoneme representations', *Journal of Memory and Language*, vol. 87, pp. 71–83, Apr. 2016, doi: 10.1016/j.jml.2015.11.002.

[8] T. Kraljic and A. G. Samuel, 'Generalization in perceptual learning for speech', *Psychonomic Bulletin & Review*, vol. 13, no. 2, pp. 262–268, Apr. 2006, doi: 10.3758/BF03193841.

[9] H. Mitterer, O. Scharenborg, and J. M. McQueen, 'Phonological abstraction without phonemes in speech perception', *Cognition*, vol. 129, no. 2, pp. 356–361, Nov. 2013, doi: 10.1016/j.cognition.2013.07.011.

[10] A. G. Samuel, 'Psycholinguists should resist the allure of linguistic units as perceptual units', *Journal of memory and language*, vol. 111, pp. 104070-, 2020, doi: 10.1016/j.jml.2019.104070.

[11] H. Mitterer, E. Reinisch, and J. M. McQueen, 'Allophones, not phonemes in spoken-word recognition', *Journal of Memory and Language*, vol. 98, pp. 77–92, Feb. 2018, doi: 10.1016/j.jml.2017.09.005.

[12] N. Kazanina, J. S. Bowers, and W. Idsardi, 'Phonemes: Lexical access and beyond', *Psychonomic bulletin & review*, vol. 25, no. 2, pp. 560–585, 2018, doi: 10.3758/s13423-017-1362-0.

[13] E. Reinisch, D. R. Wozny, H. Mitterer, and L. L. Holt, 'Phonetic category recalibration: What are the categories?', *Journal of phonetics*, vol. 45, p. 91, Jul. 2014, doi: 10.1016/j.wocn.2014.04.002.

[14] C. Kurumada, M. Brown, and M. K. Tanenhaus, 'Effects of distributional information on categorization of prosodic contours', *Psychonomic bulletin & review*, vol. 25, no. 3, pp. 1153–1160, 2017, doi: 10.3758/s13423-017-1332-6.

[15] H. R. Bosker, 'Evidence For Selective Adaptation and Recalibration in the Perception of Lexical Stress', *Language and speech*, vol. 65, no. 2, pp. 472–490, 2022, doi: 10.1177/00238309211030307.

[16] M. J. W. Yip, *Tone*. in Cambridge textbooks in linguistics. Cambridge: University Press, 2002.

[17] J. Yuan and Y. Chen, '3 rd tone sandhi in Standard Chinese: A corpus approach', *Journal of Chinese Linguistics*, vol. 42, Jan. 2014.

[18] W. S.-Y. Wang and K.-P. Li, 'Tone 3 in Pekinese', *Journal of Speech and Hearing Research*, vol. 10, no. 3, pp. 629–636, Sep. 1967, doi: 10.1044/jshr.1003.629.

[19] X. Li and Y. Chen, 'Representation and Processing of Lexical Tone and Tonal Variants: Evidence from the Mismatch Negativity', *PLoS ONE*, vol. 10, no. 12, pp. e0143097-, 2015.

[20] J. S. Nixon, Y. Chen, and N. O. Schiller, 'Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones', *Language, Cognition and Neuroscience*, vol. 30, no. 5, pp. 491–505, May 2015, doi: 10.1080/23273798.2014.942326.

[21] Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.12, retrieved 2 May 2022 from http://www.praat.org/.

[22] A. L. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed, 'Gorilla in our midst: An online behavioral experiment builder', *Behav Res Methods*, vol. 52, no. 1, pp. 388–407, Feb. 2020, doi: 10.3758/s13428-019-01237-x.

[23] R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[24] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, vol. 67, pp. 1–48, Oct. 2015, doi: 10.18637/jss.v067.i01.

[25] Cesko C. Voeten (2014). buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. R package version 2.11. https://CRAN.R-project.org/package=buildmer.

[26] Russell V. Lenth, Ben Bolker, Paul Buerkner, Iago Giné-Vázquez, Maxime Herve, Maarten Jung, Jonathon Love , Fernando Miguez, Hannes Riebl, Henrik Singmann (2014). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.0. https://CRAN.R-project.org/package=emmeans.

[27] S. Holm, 'A Simple Sequentially Rejective Multiple Test Procedure', *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[28] J. M. McQueen and L. Dilley, 'Prosody and Spoken-Word Recognition', in *The Oxford Handbook of Language Prosody*, C. Gussenhhoven and A. Chen, Eds., Oxford University Press, 2020, p. 0. doi: 10.1093/oxfordhb/9780198832232.013.33.

[29] M. Scott, 'Interaural recalibration of phonetic categories', *The Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. EL164–EL170, Feb. 2020, doi: 10.1121/10.0000735.

[30] J. Charoy, 'Accommodation to Non-Native Accented Speech: Is Perceptual Recalibration Involved?', Ph.D. dissertation, State University of New York at Stony Brook, New York, 2021. [Online]. Available: http://www.proquest.com/docview/2562832403/abstract/4ACE182BE74C457FPQ/1.