Mental Representation of Mandarin Tone 3: an Integrated Phonetic and Phonological Reflection

Ye Yanyuan, Peng Gang
The Hong Kong Polytechnic University

Abstract

The phonetic description and phonological analysis of Mandarin Tone 3 has been a complex issue that attracted divergent views. To better understand this tone, the current study has computed the mental representation of Tone 3 by adopting the reverse-correlation paradigm. Thirty participants (15 males, mean age = 21.94 ± 2.4 years) were recruited to compare and judge which of the two randomly generated stimuli sounded more like Tone 3. Analyses on the interaction between participants' response and the manipulated random contour in perception of this tone has indicated that mental representation of Mandarin Tone 3 reflected some of the phonological representations (i.e., the [+low] feature) as well as preserving the phonetic characteristics (i.e., contourisity, dynamic and static portions, and duration-dependency). This method has offered possibilities to better understand the nature of linguistic elements in an integrated way.

1 Introduction

For citation forms in the Mandarin tonal system, Tone 3 is uniquely characterized as the only concave tone. Traditionally, it has been described as having a low-falling pitch at the beginning, followed by a rising pitch towards the end of the contour, and is thus analyzed as 214 on a five-scale system where numerical sequences represent pitch values, as proposed by Chao (1965, 2013).

However, Tone 3 exhibits significant variance in its phonetic realization, especially in continuous speech, where the final rising portion may be omitted, resulting in a low-falling tone (Ho, 1976; Howie, 1974; Zhu, 2012). Additionally, the pitch contour of Tone 3 varies with different syllable durations. The concave contour is predominantly observed in longer syllables, while shorter syllables tend to exhibit low-falling variants (Howie, 1974; Nordenhake & Svantesson, 1983; Yang et al., 2017).

In the auditory domain, perceptual tasks on lexical tones have shown that Tone 3 exhibits considerable within-category variation compared to the other three tones. Despite occupying a relatively large perceptual space, the highest identification rate for Tone 3 is reported for the low-falling contour, even when stimuli are presented in isolation, as shown in Figure 1 (Peng et al., 2012). Additionally, low-level tones, although less frequently observed in production, could also be perceived as Tone 3 (Whalen & Xu, 2009). The effect of syllable duration on tone identification has also been observed, with longer durations eliciting more Tone 3 responses rather than Tone 2 (Blicher et al., 1990).

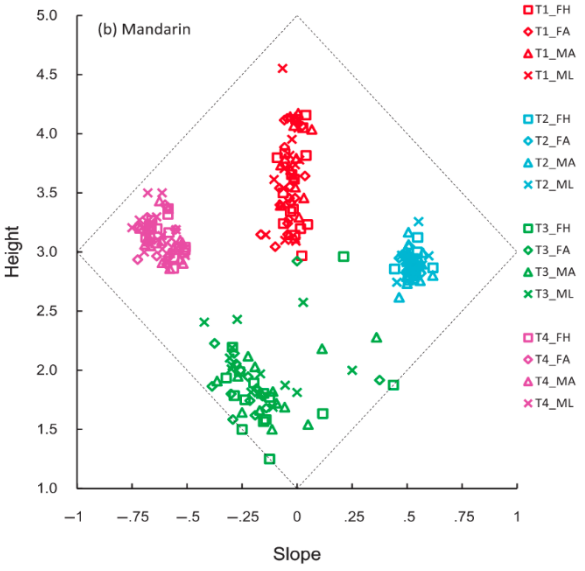

Figure 1 Two-dimensional plot (height and slope) of perceptual center of gravity for the Mandarin tone system, sourced from Peng, et al. (2012).

Phonologically, Tone 3 is generally described as having a [+low] feature, starting from early generative phonology using single-tier analysis. Milliken (1989, as cited in Duanmu, 2007) argued that the L/Low feature is crucial for Tone 3, with the final rising (H/High) being a floating feature. Duanmu, (2007) suggested that only the L feature should be considered, with the rising portion being a product of the disyllabic foot, which should be excluded. Despite their differences, both analyses emphasized the significance of the [+Low] feature and avoided focusing on the pitch shape. In multi-tiered representations, Yip (1980) proposed that the phonological representation of Tone 3 should not only focus on the [-upper] register, which she agreed is the most important aspect, but also include the contour feature, represented as LL, indicating a level and unchanging pitch direction (Bao, 1999). The view that Mandarin Tone 3 is a low-level tone was also supported by Shi & Ran (2011). These studies emphasized the dominance of the register in analyzing Tone 3, but also acknowledged the importance of the contour or pitch direction, although their views on the pitch shape varied.

The discrepancies between phonetic and phonological analyses, as well as within each domain, have reflected different understandings of Tone 3, leading to difficulties in experimental design. Based on their divergent understandings, even experiments addressing the same issue of Tone 3 have used different stimuli. For example, studies considering Tone 3 as a concave tone tended to manipulate the turning point of the pitch contour (Liu, 2004), those viewing it as low-falling focused on the pitch slope, and those treating Tone 3 as a low-level tone examined the effect of the register (Chen et al., 2010; Wang et al., 2014). These differing criteria for stimulus selection have resulted in varied outcomes, which in turn hindered the better understanding of Tone 3.

To better describe, analyze, and understand Tone 3 (as well as other tones), the reverse-correlation paradigm, a new data-driven method, was adopted in the current study. This paradigm focuses on how top-down representations are used to process incoming stimuli and can estimate the mental representations of categories based on participants' response patterns to randomly varying information (Brinkman et al., 2019). Initially applied in the visual domain, this paradigm has been adapted to auditory research with the development of the CLEESE toolbox (Burred et al., 2019). CLEESE is a Python-based toolbox that enables random and numerous manipulations of pitch on the same base audio. Several studies have begun using this novel method in auditory perception. Wang et al. (2022) found that the mental representation of typically developing children is similar but less variable than that of children with autism spectrum disorder by generating speech, complex tone, and song stimuli with randomly manipulated pitch contours.

The current study explores the mental representation of Mandarin tones through this novel paradigm, examining the relationship between phonetic realizations, phonological analyses, and mental representations under the case of Tone 3 description. The reverse-correlation paradigm will be applied to compute the mental representations. The study also seeks to examine the effect of duration on the mental representation of the tone. Thus, long and short stimuli will be generated, and their induced patterns will be compared. As the primary cue of lexical tones is fundamental frequency (f0), which becomes the main parameter used to manipulate tone variation, we also aim to determine whether listeners' pitch sensitivity affects the mental representation of the tone. Therefore, a pitch-judgment task will be conducted to examine the ability to discriminate pitch height, and the correlation between pitch sensitivity and mental representation will be analyzed. These efforts aim to better describe and understand Mandarin Tone 3, thereby investigating the nature of speech.

2 Methods

2.1 Participants

Thirty native Mandarin speakers (15 males, mean age = 21.94 ± 2.4 years) were recruited for this study. All participants were right-handed and had no background in music, linguistics, or psychology. None reported any hearing or mental health issues.

2.2 Stimuli

The Mandarin monosyllable *yi* /i/ with Tone 3 was produced by a female native Mandarin speaker who was born and grew in Beijing. A recording sample without creaky voice was selected, as this phonation type, although common observed in Tone 3 production, can disrupt the continuity of the extracted f0 contour. The pitch contour of the chosen sample was flattened to 188 Hz and

normalized to 80 dB. To investigate the effect of the duration of syllable, the sample was adjusted to 250 ms for the short condition and 450 ms for the long condition, creating the base stimuli. Then, pitch-shifting manipulation using Gaussian pitch noise was applied to these stimuli by sampling pitch values at 7 equal and successive time points, using a normal distribution (SD = 180 cents) in the CLEESE toolbox, following Burred et al. (2019). CLEESE is a Python-based toolbox used to generate variations of sound with randomly manipulated pitch contours while maintaining constant amplitude and duration, as shown in Figure 2.

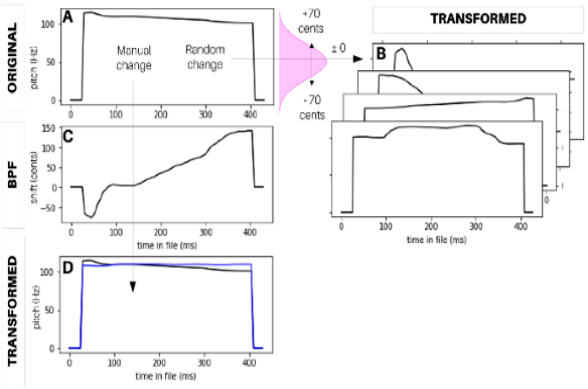

Figure 2 Examples of pitch manipulations created with CLEESE, sourced from Burred et al. (2019).

We optimized the algorithm by adding a condition that the manipulated pitch contour must change gradually, rather than abruptly or with large fluctuations, as restricted by human vocalization. After piloting the quality and naturalness of the generated syllables and the duration of the task, 200 pairs for the long condition and 200 pairs for the short condition were generated, resulting in 800 pitch variations in total.

The stimuli for the pitch-judgment task were generated as pure tones in Praat (Boersma & Weenink, 2009) with f0 set to be level and ranging from 440 to 452 Hz, resulting in 12 stimuli in total. The duration was set to 1 second and the intensity to 80 dB.

2.3 Procedure

The experiment was conducted via Gorilla, an online platform validated for effective remote experimentation (Anwyl-Irvine et al., 2020). Participants were instructed to complete two tasks: the word-comparison task and the pitch-judgment task. For the first task, aimed at computing the mental representation of Tone 3, participants were asked to choose which syllable sounded more like the word *yi3* ("以") from two stimuli randomly generated by CLEESE. There were two separate blocks for long and short durations, with the order of these blocks counterbalanced across participants.

For the pitch-judgment task, participants listened to two pure tones with pitch differences ranging from 1 Hz to 12 Hz and judged whether the second stimulus sounded higher, lower, or the same in pitch compared to the first one.

2.4 Analysis

The auditory classification image for each participant was generated using the CLEESE toolbox by averaging the differences reflected in each choice made by the participants, as described in Burred et al. (2019). To analyze the overall contouring pattern, a one-way ANOVA was conducted to compare the normalized pitch in Tone 3's mental representation at each timepoint of the pitch contour, determining whether the mental representation of Tone 3 is a level tone. Subsequently, a linear mixed-effects model was constructed to examine the effects of duration of the syllable (long vs. short) and listeners' pitch sensitivity on the normalized pitch at each timepoint using the *lme4* package in R (Bates et al., 2015).

3 Results

For the overall contour of normalized pitch in the mental representation of Tone 3, the one-way ANOVA revealed a significant main effect of Timepoint ($F(2.001) = 67.16$, $p < 0.001$). Post-hoc analysis indicated that the normalized pitch of all adjacent timepoints (i.e., Timepoint 1 vs. 2; 2 vs. 3; …; 6 vs. 7) were significantly different, as shown in Figure 3. However, no significant differences

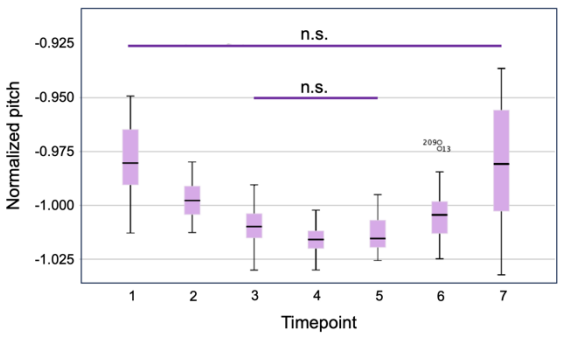

Figure 3 Normalized pitch of mental representation of Tone 3 in each timepoint

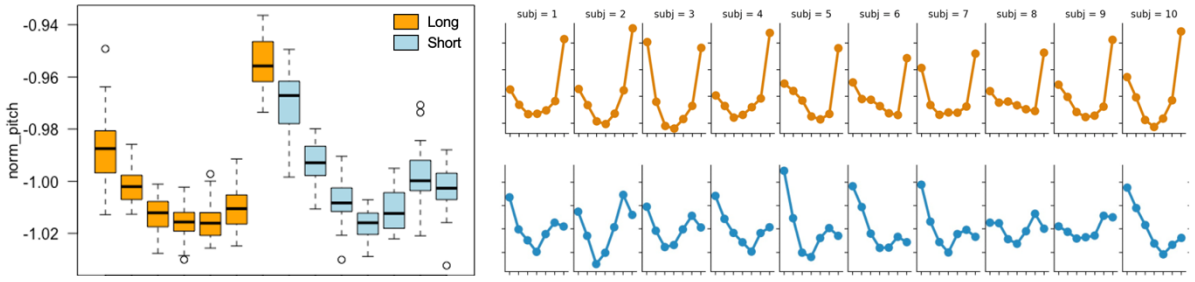

Figure 4 Mental representation of Tone 3 in long and short duration in the group (left panel) and individual level (right panel). The upper right panel (orange) shows the mental representation of long syllable from subject 1-10, the bottom right panel (blue) shows the representation of short syllable from the same subjects.

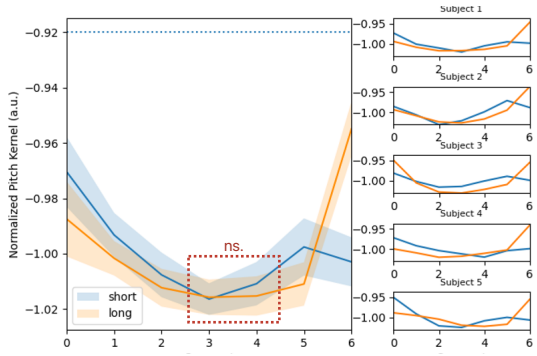

Figure 5 Classification image of Tone 3 generated by CLEESE

Time	estimate	df	*t* ratio	*p* value
T1	0.017	364	7.343	<.0001
T2	0.008	364	3.626	0.0003
T3	0.005	364	1.998	0.0465
T6	0.013	364	5.759	<.0001
T7	-0.048	364	-20.672	<.0001

Table 1 Pairwise comparison of long and short duration in each timepoint

were found between Timepoint 1 vs. 7 and Timepoint 3 vs. 5 (*p*s > 0.05).

Regarding the effects of syllable duration and listeners' pitch sensitivity, the linear mixed-effects model on the normalized pitch at each timepoint, with Duration (long vs. short) and Pitch sensitivity of listeners as fixed effects, revealed no significant fixed effect of pitch sensitivity, nor any relevant interaction effects (all *p*s > 0.05). Consequently, pitch sensitivity was removed from the model and excluded from further analysis.

A two-way repeated measures ANOVA with Duration (long vs. short) and Timepoint (1, 2, …, 7) revealed significant main effects of Duration ($F(1,29) = 30.42$, $p < 0.001$), Timepoint ($F(2.10, 60.83) = 113.90$, $p < 0.001$), and their interaction ($F(2.48, 71.86) = 117.88$, $p < 0.001$). Post-hoc analysis showed that at Timepoint 1, 2, 3, and 6, the normalized pitch of the short syllables was higher than that of the long syllables. Conversely, at Timepoint 7, the normalized pitch of the short syllable was lower than that of the long syllable (see Table 1). At Timepoints 4 and 5, the differences in normalized pitch between long and short syllables did not reach a significant level (*p*s > 0.05). The generated classification image of Tone 3 was shown in Figure 5.

4 Discussions

This study seeks to better describe and analyze Mandarin Tone 3 by adopting the reverse-correlation paradigm using CLEESE toolbox. The comparison between the computed mental representation and the phonetic descriptions and phonological analyses are conducted based on the observation of their overall shape of pitch contour and the effects elicited by duration. The [+low] pitch is the most crucial feature in the mental representation of Mandarin Tone 3, as supported by the results of both analyses. In the overall pitch contour, Timepoint 4 exhibited the lowest pitch height with the smallest variance compared to the other six timepoints. Additionally, there is a consistently low portion in the middle part of the pitch contour, regardless of the duration of syllables. Specifically, when the duration is longer, the onset of the syllable becomes lower, and the offset becomes higher than that of the shorter syllable. Notably, both long and short syllables have shared a portion that represents the lowest part of the contour.

The significance of low pitch in Tone 3's mental contour reflects its phonetic descriptions. Although there are differing views on the shape of its pitch — whether low-falling or concave — they all indicate that it should be low at one portion (Zhu et al., 2012). The crucial role of the [+low] feature also aligns with the phonological analysis of Tone

3, which emphasizes its [low] or [-upper] feature (Duanmu, 2007; Yip, 1980). Notably, the phonological view mainly emphasizes that [+low] is the most important feature, denying the role of contour in the analysis of Tone 3, which has also been addressed in the analysis of mental representation.

For the overall contour, as indicated by the result that the normalized pitch of all adjacent timepoints were significantly different, the mental representation of Mandarin Tone 3 is concave rather than level. Specifically, the normalized pitch at the first and last timepoints of the contour are the highest, while the middle part (i.e., Timepoint 4) has the lowest pitch height. Although this shape of the contour would be altered by duration, that longer syllables have a relatively lower onset and higher offset, it remains concave in both long and short syllables.

This finding contrasts with the idea that the phonological representation of Tone 3 is low-level (Yip, 1980; Shi & Ran, 2011) and shows consistency with the phonetic description. The reason that Tone 3 is phonologically analyzed as a level tone is mainly rooted in systematic rules and abstraction rather than merely focusing on the acoustic aspect. For instance, the view that Tone 3 should be considered as a level tone is based on the perspective that when level pitch is acceptable to be a phonetic variant, it could be considered the underlying form (Maddieson, 1977). The inconsistency between the mental and phonological representation may reflect the distinction of focus, such that the mental representation might be influenced and thus focuses more on the acoustic characteristics.

As indicated by the results of the current study, to figure out the characteristics of Mandarin Tone 3, [+low] is indeed crucial. But contour is inevitable in the description as well. It might not matter about the exact shape of the concave, but [+contourisity] is significant (proposed by Zhu 2012). In phonetic realization, the shape of the concave might be a result of the adjustment of gestures approaching to the [+low] target and thus would be affected by the duration of the syllable to be produced. This "Durational effect" has also been observed in the perceptual task, indicating a possible effect elicited by production (Blicher et al., 1990). In the mental representation of Tone 3, we have also observed the concave pitch in both long and short tones, meaning that the seemly irrelevant adjustment of vocalization does contribute to how we define and recognize a tone.

The larger pitch variance at the onset and offset, particularly at Timepoints 1 and 7, suggests the presence of two dynamic portions at the beginning and ending of the contour in Tone 3. Conversely, a static portion in the middle, as indicated by Timepoint 4 where pitch variance is relatively low, reflects the consistent occurrence of the low pitch in the contour.

This perspective that Tone 3 may encompass both dynamic and static portions is further supported by the involvement of varying syllable durations. Our findings indicate that longer syllables exhibit a lower onset and higher offset, whereas shorter syllables display the opposite pattern. However, there was no significant difference between long and short syllables in the middle portion of the contour. This has suggested that the change in duration does not affect the pitch in this middle portion, which remains static and is resistant to be changed. This finding is in line with Zhang & Shi (2016), who have described that there are dynamic and static portions in the pitch contour of Mandarin Tone 3 based on the large sample size of Beijing Mandarin vocalizations. The dynamic and static characteristics observed in both phonetic and mental representation of Tone 3 have reflected a correspondence between acoustic and cognitive domains.

In the mixed-effects model, we failed to observe the fixed effect of participants' pitch sensitivity or the interaction effects with other factors on their mental representation of Tone 3, suggesting that auditory sensitivity might not contribute to the pattern of the mental representation of lexical tones focusing on the same acoustic parameters.

5 Conclusions

Based on the computation and observation upon listeners response pattern to the random manipulated f0 contours in the reverse-correlation paradigm, mental representation of Mandarin Tone 3 is found to have reflected some of the phonological representations (i.e., [+low]) as well as preserving the phonetic characteristics (i.e., contourisity, dynamic and static portions, and duration-dependency). By reflecting a combination of phonetic and phonological characteristics, this method has offered possibilities to better understand the nature of linguistic elements in an integrated way.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.

Bao, Z. (1999). *The Structure of Tone*. Oxford University Press, USA.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37–49. https://doi.org/10.1016/S0095-4470(19)30357-2

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer (Version 5.1. 05)[Computer program]. Retrieved May 1, 2009*.

Brinkman, L., Goffin, S., Schoot, R., Haren, N. E. M., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, *51*. https://doi.org/10.3758/s13428-019-01232-2

Burred, J. J., Ponsot, E., Goupil, L., Liuni, M., & Aucouturier, J.-J. (2019). CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. *PLOS ONE*, *14*(4), e0205943. https://doi.org/10.1371/journal.pone.0205943

Chao, Y. R. (1965). *A GRAMMAR OF SPOKEN CHINESE*.

Chao, Y. R. (2013). Mandarin primer. In *Mandarin Primer*. Harvard University Press.

Chen, X. D. (2010). The perceptual boundary between Mandarin Yinping and Shangsheng tones. In *Proceedings of the 9th Conference on Chinese Phonetics* (pp. 6). Chinese Phonetics Society, Chinese Acoustics Society, Language, Music and Hearing Committee, Chinese Information Processing Society of China.

Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Ho, A. T. (1976). The Acoustic Variation of Mandarin Tones. *Phonetica*, *33*(5), 353–367. https://doi.org/10.1159/000259792

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30*(3), 129–148.

Liu, J. (2004). Perceiving the boundary between the lexical rising tone and the falling-rising tone. In *Festschrift for Professor Wang Shiyuan's 70th Birthday* (pp. 222-233). Tianjin: Nankai University Press.

Maddieson, I. (1977). *Universals of Tone: Six Studies*. University of California.

Nordenhake, M., & Svantesson, J. O. (1983). Duration of standard Chinese word tones in different sentence environments. Working Papers/Lund University, Department of Linguistics and Phonetics, 25.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., & Wang, W. S.-Y. (2012). The Effect of Intertalker Variations on Acoustic–Perceptual Mapping in Cantonese and Mandarin Tone Systems. *Journal of Speech, Language, and Hearing Research*, *55*(2), 579–595. https://doi.org/10.1044/1092-4388(2011/11-0025)

Shi, F., & Ran, Q. B. (2011). The essence of Mandarin Shangsheng is a low-level tone: A reanalysis of "The perception of level tones in Chinese". *Chinese Language,* (6), 550-555.

Wang, L., Ong, J. H., Ponsot, E., Hou, Q., Jiang, C., & Liu, F. (2022). Mental representations of speech and musical pitch contours reveal a diversity of profiles in autism spectrum disorder. *Autism*, 13623613221111207. https://doi.org/10.1177/13623613221111207

Wang, P., Shi, F., Rong, R., Chen, X. D., Li, S., & Wang, X. X. (2014). The perceptual category of Mandarin Shangsheng. *Chinese Language*, (04), 359-370+384.

Whalen, D. H., & Xu, Y. (2009). Information for Mandarin Tones in the Amplitude Contour and in Brief Segments. *Phonetica*, *49*(1), 25–47. https://doi.org/10.1159/000261901

Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the Duration of Mandarin Tones. *Interspeech 2017*, 1407–1411. https://doi.org/10.21437/Interspeech.2017-29

Yip, M. J. (1980). *The tonal phonology of Chinese* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/15971

Zhang, Y., & Shi, F. (2016). Statistical analysis of Mandarin single-syllable tones. *Chinese Journal of Phonetics*, (00).

Zhu, X., Yi, L., & Zhang, T. (2012). Types of dipping tones. In *Tonal Aspects of Languages-Third International Symposium*.