CrossMark

1799

# Encoding lexical tones in jTRACE: a simulation of monosyllabic spoken word recognition in Mandarin Chinese

Lan Shuai[1] · Jeffrey G. Malins[1]

**Abstract** Despite its prevalence as one of the most highly influential models of spoken word recognition, the TRACE model has yet to be extended to consider tonal languages such as Mandarin Chinese. A key reason for this is that the model in its current state does not encode lexical tone. In this report, we present a modified version of the jTRACE model in which we borrowed on its existing architecture to code for Mandarin phonemes and tones. Units are coded in a way that is meant to capture the similarity in timing of access to vowel and tone information that has been observed in previous studies of Mandarin spoken word recognition. We validated the model by first simulating a recent experiment that had used the visual world paradigm to investigate how native Mandarin speakers process monosyllabic Mandarin words (Malins & Joanisse, 2010). We then subsequently simulated two psycholinguistic phenomena: (1) differences in the timing of resolution of tonal contrast pairs, and (2) the interaction between syllable frequency and tonal probability. In all cases, the model gave rise to results comparable to those of published data with human subjects, suggesting that it is a viable working model of spoken word recognition in Mandarin. It is our hope that this tool will be of use to practitioners studying the psycholinguistics of Mandarin Chinese and will help inspire similar models for other tonal languages, such as Cantonese and Thai.

Since its publication in 1986, McClelland and Elman's TRACE model has remained one of the most highly influential computational models of spoken word recognition. This model—and refutations of it—has helped shape psycholinguistic research over the past 30 years by allowing researchers to test specific hypotheses and to generate new ones (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Marslen-Wilson & Warren, 1994; Protopapas, 1999). In recent years, this powerful tool has become even more accessible to researchers in the community, thanks to the work of Strauss, Harris, and Magnuson (2007), who released an implementation of the TRACE model using a Java platform (jTRACE). This implementation has since been used in a number of studies spanning various domains, including speech and language disorders and infant word recognition (e.g., Mayor & Plunkett, 2014; McMurray, Samelson, Lee, & Tomblin, 2010; Mirman, Yee, Blumstein, & Magnuson, 2011).

Despite this increased accessibility, TRACE simulations have yet to be conducted in many of the world's widely spoken languages, including Mandarin Chinese (Malins & Joanisse, 2012a). A key reason for this is that TRACE does not currently encode lexical tone, which is a fundamental feature of tonal languages such as Mandarin. This represents a considerable roadblock in the development of the emerging field of the psycholinguistics of East Asian languages, many of which are tonal in nature (Li, Tan, Bates, & Tzeng, 2006; Zhou & Shu, 2011).

Lan Shuai and Jeffrey G. Malins contributed equally to this study and should be considered co-first authors.

✉ Lan Shuai
shuai@haskins.yale.edu

✉ Jeffrey G. Malins
malins@haskins.yale.edu

[1] Haskins Laboratories, 300 George Street Suite 900, New Haven, CT 06511, USA

Springer

*Lexical tone* refers to variation in the fundamental frequency of a speaker's voice used to differentiate the meanings of spoken words. For example, in Mandarin, the word *hua* can refer to "flower" when pronounced in a high and level tone, whereas it can refer to "painting" when pronounced in a falling tone. As this example illustrates, processing the individual consonants and vowels—or phonemes—in a word is not enough for a listener to access the meaning of a spoken word; tonal information must be accessed as well. Previous work by several groups has suggested that tonal information is accessed online during the unfolding of a spoken word over a similar time course as the vowels upon which tones are carried (Brown-Schmidt & Canseco-Gonzalez, 2004; Malins et al., 2014; Malins & Joanisse, 2010, 2012a; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Zhao, Guo, Zhou, & Shu, 2011). Therefore, it is vitally important that a model of spoken word recognition in Mandarin be developed with this finding in mind. As has been suggested previously, continuous mapping models such as TRACE are well-suited to capture the dynamic nature of spoken word processing in tonal languages such as Mandarin and Cantonese (Malins & Joanisse, 2012b; Tong, McBride, & Burnham, 2014; Ye & Connine, 1999; Zhao et al., 2011).

In this article, we present a computational model called TRACE-T, which we designed to simulate the recognition of monosyllabic spoken words in Mandarin Chinese. To develop this model, we modified the existing jTRACE architecture to code for Mandarin phonemes and tones. Spoken words are represented using a coding scheme that allows for similar timing of access to tonal and vowel information. Furthermore, we validated this model by simulating published eyetracking data to show that this model simulates Mandarin spoken word recognition in a way that captures the patterns observed in experimental data from human subjects. It is our hope that the community can make use of this model to move forward psycholinguistic research in Mandarin Chinese and other tonal languages.

## How the TRACE-T model works

The TRACE model is a connectionist model with three layers, corresponding to different grain sizes of spoken words. At the highest level of the model are units corresponding to single words, or lexical representations. Below this is a level containing the phonemic units that make up spoken words. At the lowest level of the model are the features associated with different phonemes, such as power and voicing. Information flows through the model in a bidirectional fashion, with excitatory connections between levels of the model. Within layers, there are lateral inhibitory connections. Together, these excitatory and inhibitory connections affect the extent of activation of different units in the model in response to a given input. Items sharing phonological similarity, such as words that share similar phonemes, or phonemes that share similar

features, compete for activation as temporal slices of the input are presented to the model.

To develop the TRACE-T model, we first needed to code Mandarin segmental structure. Instead of the seven feature dimensions in jTRACE—namely, *consonantal*, *vocalic*, *diffuseness*, *acuteness*, *voicing*, *power*, and *burst amplitude*—we chose instead to categorize Mandarin phonemes along dimensions more appropriate for the Mandarin phonemic inventory. We elected to employ the dimensions used in the Mandarin Chinese version of PatPho, an online database of phonological representations of Mandarin syllables (Zhao & Li, 2009). In the Mandarin version of PatPho, phonemes are classified along three dimensions, each of which lump together two different types of features—one for consonants and one for vowels. We took each of these dimensions and split them apart into two dimensions, giving rise to six dimensions in total: *voicing*, *place of articulation*, and *manner of articulation* for consonants, and *roundedness*, *tongue position*, and *tongue height* for vowels. We selected an appropriate level (out of eight levels in total) for each of these dimensions, and then assigned binary values to these levels. As in the original TRACE model (McClelland & Elman, 1986), there was also a silent phoneme unit that was assigned

**Table 1** Coding scheme employed for Mandarin segmental units in TRACE-T

| IPA | Pinyin[a] | Coding Scheme[b] | IPA | Pinyin | Coding Scheme |
|---|---|---|---|---|---|
| /ɑ/ | [a] *in* [ang] | 8–0–1–0–1–0 | /kʰ/ | [k] | 0–3–0–1–0–6 |
| /a/ | [a] | 8–0–3–0–1–0 | /n/ | [n] | 0–8–0–5–0–8 |
| /ɛ/ | [e] *in* [ie] | 8–0–6–0–3–0 | /m/ | [m] | 0–8–0–8–0–8 |
| /e/ | [e] *in* [ei] | 8–0–6–0–5–0 | /ŋ/ | [ng] | 0–8–0–1–0–8 |
| /o/ | [o] | 1–0–1–0–7–0 | /l/ | [l] | 0–8–0–5–0–1 |
| /ɤ/ | [e] | 8–0–1–0–7–0 | /s/ | [s] | 0–1–0–5–0–4 |
| /ə/ | [e] *in* [en] | 8–0–3–0–7–0 | /f/ | [f] | 0–1–0–7–0–4 |
| /u/ | [u] | 1–0–1–0–8–0 | /ɕ/ | [x] | 0–1–0–2–0–4 |
| /y/ | [ü] | 1–0–6–0–8–0 | /zʅ/ | [r] | 0–8–0–4–0–4 |
| /i/ | [i] | 8–0–6–0–8–0 | /sʅ/ | [sh] | 0–1–0–4–0–4 |
| /ɿ/ | [i] *in* [zi] | 8–0–8–0–8–0 | /x/ | [h] | 0–1–0–1–0–4 |
| /ʅ/ | [i] *in* [zhi] | 8–0–7–0–8–0 | /ts/ | [z] | 0–1–0–5–0–3 |
| /t/ | [d] | 0–1–0–5–0–6 | /tsʰ/ | [c] | 0–3–0–5–0–3 |
| /tʰ/ | [t] | 0–3–0–5–0–6 | /tɕ/ | [j] | 0–1–0–2–0–3 |
| /p/ | [b] | 0–1–0–8–0–6 | /tɕʰ/ | [q] | 0–3–0–2–0–3 |
| /pʰ/ | [p] | 0–3–0–8–0–6 | /tʂ/ | [zh] | 0–1–0–4–0–3 |
| /k/ | [g] | 0–1–0–1–0–6 | /tʂʰ/ | [ch] | 0–3–0–4–0–3 |

IPA — the International Phonetic Alphabet. [a] Note that the same symbol in Pinyin can be used to represent two or more distinct vowels in IPA; in these instances, example Pinyin sequences containing these vowels are offered for clarification. [b] The coding scheme for each unit is given as the level (out of eight) that was assigned a binary value for each of the six respective dimensions. Values for these dimensions are listed in the following order: *vowel roundedness*, *voicing* for consonants, *tongue position* for vowels, *place of articulation* for consonants, *tongue height* for vowels, and *manner of articulation* for consonants. For example, the vowel /ɑ/ was assigned a value of 1 in the eighth level of the vowel roundedness feature, a value of 1 in the first level of the tongue position feature, and a value of 1 in the first level of the tongue height feature

**Table 2** Tonal coding scheme employed in TRACE-T

| Tone | IPA | Pinyin (Above Vowel) | Tone Units[a] | Pitch Height Coding[b] | Pitch Slope Coding[b] | Silent |
|---|---|---|---|---|---|---|
| 1 | /˥/ | [ē] | $T_{4L}$ $T_{4L}$ $T_{4L}$ $T_{4L}$ $T_{4L}$ | 2 2 2 2 2 | 6 6 6 6 6 | |
| 2 | /˧˥/ | [é] | $T_{2L}$ $T_{2L}$ $T_{3R}$ $T_{4R}$ $T_{4L}$ | 4 4 3 2 2 | 6 6 7 7 6 | |
| 3 | /˨˩˦/ | [ě] | $T_{2L}$ $T_{1F}$ $T_{1L}$ $T_{1L}$ $T_{2R}$ | 4 5 5 5 4 | 6 8 6 6 7 | |
| 4 | /˥˩/ | [è] | $T_{5L}$ $T_{5L}$ $T_{4F}$ $T_{3F}$ $T_{2F}$ | 1 1 2 3 4 | 6 6 8 8 8 | |
| With obstruent initial consonant[c] | | | $T_{silent}$ $T_{HS-2}$ $T_{HS-3}$ $T_{HS-4}$ $T_{HS-5}$ | | | 9 |

[a] Tone units $T_{HS-1}$ through $T_{HS-5}$ are shown for each of the four tones, in which the pitch heights (*H*) range from 1 (*lowest*) to 5 (*highest*), and pitch slopes (*S*) are coded in three levels: level (L), rising (R), and falling (F). [b] The values of pitch height and pitch slope were calculated from the average of 1,044 tokens (36 syllables uttered by 29 speakers) from the speech database reported in Zhou, Zhang, Lee, and Xu (2008). Pitch height and pitch slope were coded by assigning binary values to the feature levels listed (out of eight levels total) for the tonal dimension. [c] As we illustrate here, obstruent initial consonants had $T_{silent}$ in the $T_{HS-1}$ position

a value of 1 in the ninth feature level in all dimensions. The full segmental coding scheme is detailed in Table 1.

To incorporate Mandarin tones, we took the remaining, unused feature dimension in the model and used it to simultaneously code for two important distinguishing features of Mandarin tones: pitch height and pitch slope (Chandrasekaran, Gandour, & Krishnan, 2007; Gandour, 1984; Shuai, Gong, Ho, & Wang, 2016; Wang, 1967). Using previously published speech production data from a set of 29 native Mandarin speakers (36 syllables each, for a total of 1,044 tokens; Zhou, Zhang, Lee, & Xu, 2008), we took the full range of semitones spanning the four tones (which were time-normalized) and divided it into five levels of pitch height. Next, pitch slope was coded in three levels: *level*, *rising*, or *falling*. Fifteen tone units were then created, corresponding to the 15 unique combinations of pitch height and slope. We chose to label these using the scheme $T_{HS}$, where *H* corresponds to a level of height, from 1 to 5 (5 being the highest), and *S* to a slope (L = level, R = rising, F = falling). The normalized time series of each of the four tones was then divided into five temporal intervals, and the pitch slope and average pitch height were determined for each. Tone units were then selected as those with a value of 1 in the bin containing the average value of pitch height for a given interval and a value of 1 for the appropriate level of pitch slope for the interval. This tonal coding scheme is detailed in Table 2.

Syllables were represented with ten units corresponding to alternating phoneme and tone units using the following scheme: $P_1 T_{HS-1} P_2 T_{HS-2} P_3 T_{HS-3} P_4 T_{HS-4} P_5 T_{HS-5}$, where $P_1$ through $P_5$ are segmental units and $T_{HS-1}$ through $T_{HS-5}$ are five tone units, corresponding to the five intervals over which the syllable's tone is expressed. This alternation between phoneme and tone units was designed to allow for simultaneous access to phonemic and tonal information during the unfolding of a spoken word. For syllables with less than five phonemes, the tone-bearing vowel was repeated to allow the number of phonemes to equal five. For example, the syllable *xia1* was coded using the string {x $T_{silent}$ i $T_{4L}$ a $T_{4L}$ a $T_{4L}$ a $T_{4L}$}; the $T_{silent}$ unit will be explained shortly. The $P_1$ unit, at the beginning, was either an onset consonant or vowel; similarly, the $P_5$ unit corresponded to a vowel or to one of

the nasals /n/ or /ŋ/ ([n] and [ng] in Pinyin). $P_2$ through $P_4$ were always vowels. Since it has been shown that sonorant onset consonants can signal tonal information (Chen & Tucker, 2013; Howie, 1974), the nasals /m/ and /n/ and the glides and liquids /l/, /ʐ/, and /tɕ/ ([m], [n], [l], [r], and [j] in Pinyin) were allowed to carry tone, in addition to the subsequent vowels. This was done by setting $T_{HS-1}$ to a tone unit if $P_1$ corresponded to a vowel or sonorant consonant; otherwise, $T_{HS-1}$ was set to the "silent" tone unit ($T_{silent}$). As is shown in Table 2, the $T_{silent}$ unit was coded as having a value of 1 in the ninth level of the tone dimension (i.e., the level not used to code for either pitch height or pitch slope), and a value of 0 in all other levels.

The full model architecture is illustrated using a sample monosyllable in Fig. 1. Additionally, we coded a lexicon containing the 500 most frequent Mandarin syllables, as determined using *Modern Chinese Frequency Dictionary* (Beijing Language Institute, 1986). Together, the lexicon and phoneme inventory were stored along with default parameters in a .jt file, which is included in the supplementary materials.[1]

## Simulation of published experimental data

To validate the model, we first simulated data from an experiment investigating the time course of spoken word recognition in native speakers of Mandarin Chinese. This experiment (Malins & Joanisse, 2010) used the visual world paradigm to assess the time course over which listeners resolved competition between monosyllabic Mandarin words sharing different types of phonological relationships. In each trial, native Mandarin speakers (*N* = 17) were presented with four pictures

---

[1] Even though frequency information is included in the lexicon, all parameters have been left at the default level. Researchers interested in making use of this full lexicon would be advised to adjust the model parameters in a way that takes into account this frequency information; as we mention later in the text, Dahan et al. (2001) make a number of recommendations concerning how best to do this.
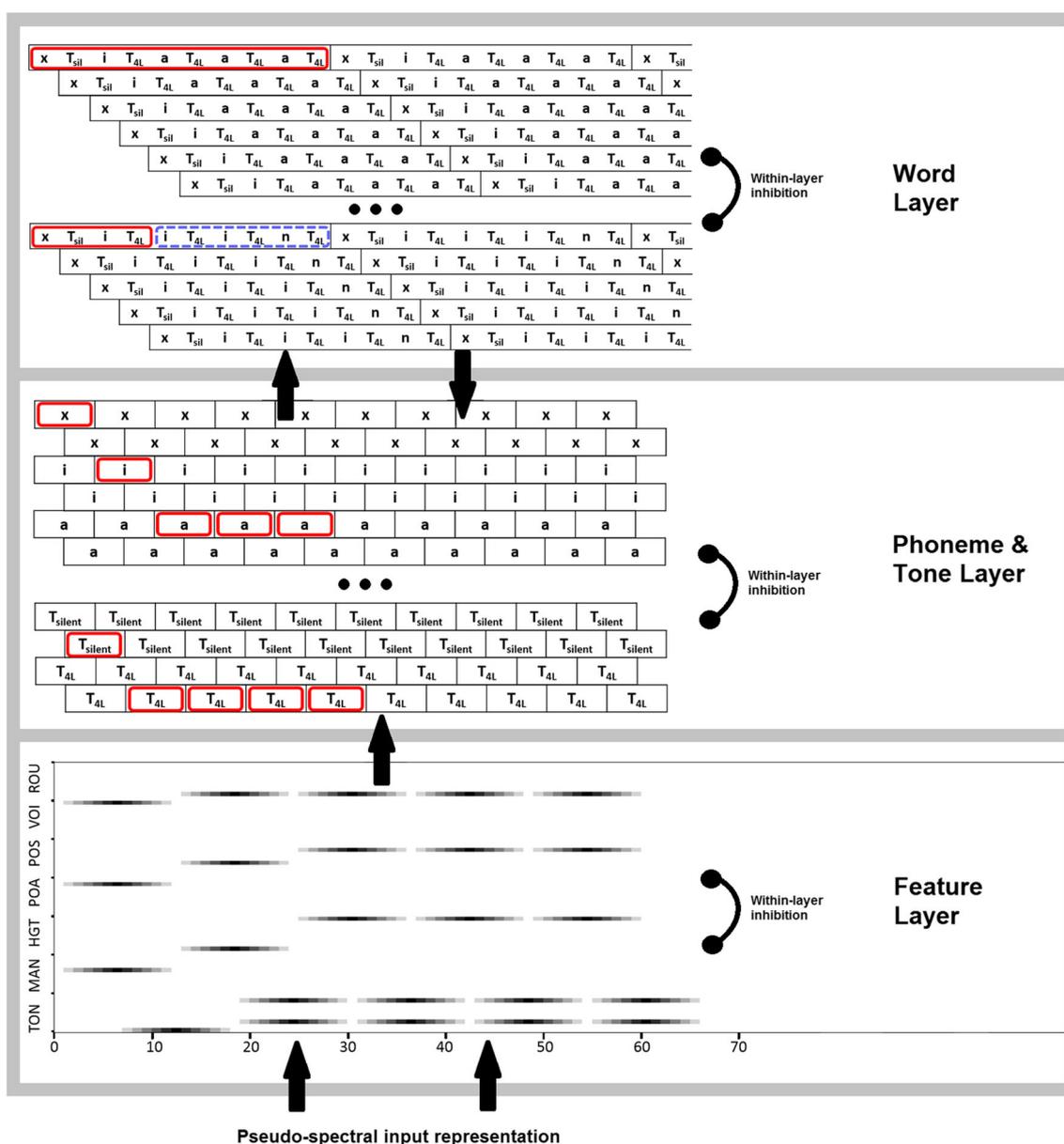
**Fig. 1** A schematic illustration of the TRACE-T model architecture. In this example, the model is presented with pseudo-spectral input corresponding to the monosyllable *xia1*, which gives rise to the activation of units in the phoneme and word layers (shown in red boxes). Monosyllables overlapping in phonemes and/or tone with the input (e.g., *xin1*, shown in the blue box) can also become partially active during the course of input presentation, giving rise to lexical competition effects. This is the type of lexical competition simulated in the following section. For the feature layer, ROU = vowel roundedness, VOI = voicing for consonants, POS = tongue position for vowels, POA = place of articulation for consonants, HGT = tongue height for vowels, and MAN = manner of articulation for consonants, TON = pitch height and pitch slope for lexical tone. The coding scheme for the phoneme and tone layers is detailed in Tables 1 and 2

in an array on a computer screen, following which they heard a spoken word matching one of the pictures. Their task was to indicate, via button press, the position of the picture in the array that matched the target word, and their eye movements were recorded while they completed this task. Importantly, in some trials the name of one other picture in the array shared phonological similarity with the target word in one or more components of the Mandarin syllable. The critical experimental manipulation thus concerned which components of the syllable were shared between targets and competitors.

More specifically, in the Malins and Joanisse (2010) experiment, the targets and competitors overlapped in segmental information but differed in tone (*segmental* competitors; e.g., *qiu2–qiu1*), overlapped in onset and tone but differed in rime (*cohort* competitors; e.g., *qiu2–qian2*), overlapped in rime and tone but differed in onset (*rhyme* competitors; e.g., *qiu2–niu2*),

**Table 3** Sets of items in each condition of the visual world paradigm experiment

| Condition | Set | Target | Competitor | Distractor | Distractor |
|---|---|---|---|---|---|
| Segmental | 1 | *chuang2* (bed) | *chuang1* (window) | *mian4* (noodles) | *xin4* (envelope) |
| | 2 | *hua1* (flower) | *hua4* (painting) | *qian2* (money) | *tui3* (leg) |
| | 3 | *mi4* (honey) | *mi3* (rice) | *chuan2* (boat) | *qiu1* (autumn) |
| | 4 | *qiu2* (ball) | *qiu1* (autumn) | *mi3* (rice) | *mian4* (noodles) |
| | 5 | *shu3* (mouse) | *shu1* (book) | *di4* (land) | *chuan2* (boat) |
| | 6 | *tu3* (dirt) | *tu4* (rabbit) | *xia1* (shrimp) | *jin1* (gold) |
| | 7 | *xin1* (heart) | *xin4* (envelope) | *gu3* (drum) | *tui3* (leg) |
| Cohort | 1 | *chuang2* (bed) | *chuan2* (boat) | *shu1* (book) | *mi3* (rice) |
| | 2 | *hua1* (flower) | *hui1* (gray) | *niu2* (cow) | *qian2* (money) |
| | 3 | *mi4* (honey) | *mian4* (noodles) | *chuang1* (window) | *qiu1* (autumn) |
| | 4 | *qiu2* (ball) | *qian2* (money) | *shui3* (water) | *hui1* (gray) |
| | 5 | *shu3* (mouse) | *shui3* (water) | *huang2* (yellow) | *chuang1* (window) |
| | 6 | *tu3* (dirt) | *tui3* (leg) | *gua1* (melon) | *xin4* (envelope) |
| | 7 | *xin1* (heart) | *xia1* (shrimp) | *tui3* (leg) | *tu4* (rabbit) |
| Rhyme | 1 | *chuang2* (bed) | *huang2* (yellow) | *xin4* (envelope) | *shu1* (book) |
| | 2 | *hua1* (flower) | *gua1* (melon) | *tu4* (rabbit) | *di4* (land) |
| | 3 | *mi4* (honey) | *di4* (land) | *qiu1* (autumn) | *chuang1* (window) |
| | 4 | *qiu2* (ball) | *niu2* (cow) | *gua1* (melon) | *hui1* (gray) |
| | 5 | *shu3* (mouse) | *gu3* (drum) | *di4* (land) | *chuan2* (boat) |
| | 6 | *tu3* (dirt) | *gu3* (drum) | *hua4* (painting) | *mian4* (noodles) |
| | 7 | *xin1* (heart) | *jin1* (gold) | *gu3* (drum) | *niu2* (cow) |
| Tonal | 1 | *chuang2* (bed) | *niu2* (cow) | *hui1* (gray) | *xia1* (shrimp) |
| | 2 | *hua1* (flower) | *jin1* (gold) | *shui3* (water) | *gu3* (drum) |
| | 3 | *mi4* (honey) | *hua4* (painting) | *shu1* (book) | *qian2* (money) |
| | 4 | *qiu2* (ball) | *huang2* (yellow) | *xia1* (shrimp) | *shui3* (water) |
| | 5 | *shu3* (mouse) | *mi3* (rice) | *huang2* (yellow) | *jin1* (gold) |
| | 6 | *tu3* (dirt) | *mi3* (rice) | *jin1* (gold) | *hua4* (painting) |
| | 7 | *xin1* (heart) | *gua1* (melon) | *tui3* (leg) | *tu4* (rabbit) |

| Condition | Set | Target | Distractor | Distractor | Distractor |
|---|---|---|---|---|---|
| Baseline | 1 | *chuang2* (bed) | *shu1* (book) | *mi4* (honey) | *xia1* (shrimp) |
| | 2 | *hua1* (flower) | *shui3* (water) | *qiu2* (ball) | *mian4* (noodles) |
| | 3 | *mi4* (honey) | *xia1* (shrimp) | *chuang2* (bed) | *shu1* (book) |
| | 4 | *qiu2* (ball) | *mian4* (noodles) | *shui3* (water) | *hua1* (flower) |
| | 5 | *shu3* (mouse) | *jin1* (gold) | *chuan2* (boat) | *qian2* (money) |
| | 6 | *tu3* (dirt) | *gua1* (melon) | *xin4* (envelope) | *hua4* (painting) |
| | 7 | *xin1* (heart) | *tu4* (rabbit) | *niu2* (cow) | *di4* (land) |

Each set was used as a four-item tuple in the simulations

or overlapped in tone only and differed in all segments (*tonal* competitors; e.g., *chuang2–niu2*).[2] In these competitor conditions, one of the pictures in the array corresponded to the target, and a second to the competitor, whereas the names of the other two pictures were phonologically unrelated to the targets and competitors in all segments and in tone. Additionally, there was a baseline control condition in which targets were presented with three phonologically unrelated distractor items (see Table 3 for the full list of experimental items).

To simulate this experiment, we used methods similar to those employed in previous studies using TRACE to simulate

visual world paradigm experiments (e.g., Allopenna et al., 1998; Dahan et al., 2001; McMurray et al., 2010). To do this, we coded each of the seven sets of items from the original experiment as "tuples" corresponding to the four pictures in the array. To conduct the simulations, a "lexicon" was constructed that contained only the 27 items used in the experiment. Thus, competition only occurred amongst items within this restricted set.

Simulations of eyetracking trials were conducted by using the Luce choice rule to convert activations to response probabilities (the constant $k$ was set to 7, following Allopenna et al., 1998). Simulations were iterated for 80 cycles, and the response probability of looking at each of the four items in the tuple was calculated for each cycle. We left all parameters at the default level except for one: stochasticity. This was set to a relatively low value of .00279 to introduce some noise in the

---

[2] Note that these condition labels have since been changed to those listed in Malins et al. (2014), to make them more consistent with other studies in the field. However, in this report we have retained the original naming system for the sake of readers who wish to compare the present set of results with those reported in the original study.

system in order to mimic the trialwise variability observed in human subjects.

In line with the Malins and Joanisse (2010) study, we only analyzed trials for which the target and competitor relationships were identical to those listed in Table 3. This constituted only half of the data, since in the other half of trials the roles of the targets and competitors were inverted, so that it was equally likely that subjects would hear targets or competitors on any given trial. As a result, we analyzed 35 trials per condition, as in the original study. Across the five blocks of the experiment, each of the seven sets of items in each condition was repeated once per block. Therefore each block was composed of the same set of trials, and blocks only differed with respect to trial order.

To compare the simulation data with the experimental data, we reanalyzed the original data by taking the grand average across subjects and items within each block of the experiment, giving rise to five data points per condition at each time point (i.e., five repetitions of the same set of trials in five different orders). To generate the same number of data points for the simulations, we ran the seven sets of tuples per condition

though the model five times, taking the average value across the tuples at each time point for each simulation. For the human data, we analyzed between 200 and 1,100 ms post-stimulus onset, as in the original experiment. Because the eyetracker recorded looks to target items at 60 Hz, this analysis window corresponded to the 13th through the 66th time point recorded. For the simulations, we also chose to analyze the time course over which looks to target began to rise and subsequently reached asymptote; this neatly corresponded to cycles 13 through 66 in the model.

We performed growth curve analyses on both the experimental and simulation data, as this technique is highly appropriate for longitudinal data such as looks to items over time (Mirman, 2014). To conduct these analyses, we used the *lme4* package in R to develop a model using fourth-order orthogonal polynomials to capture the time course of looks to target over time. This model contained fixed effects of condition on all time terms, as well as trial repetition and repetition-by-condition random effects on all time terms. The baseline control condition was used as a reference, and parameters for each
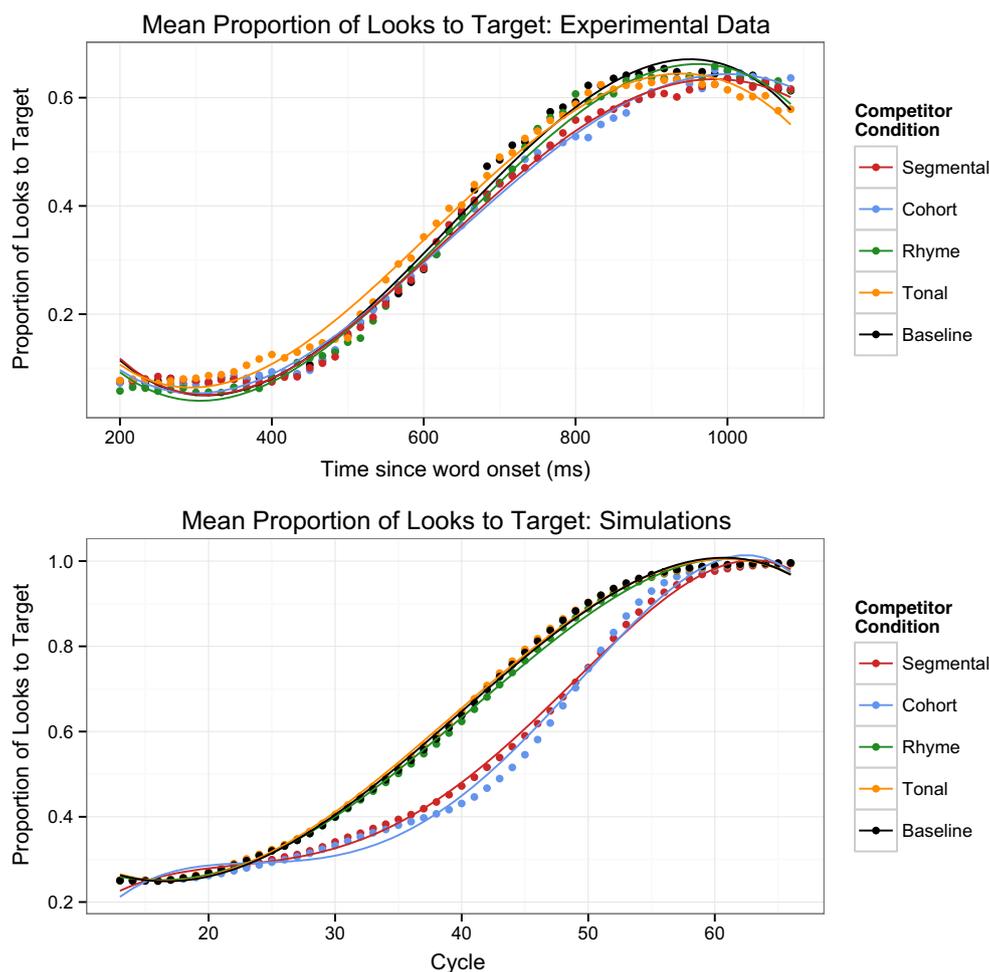


**Fig. 2** Mean proportions of looks to target in the different competitor conditions for both the experimental (top panel) and simulation (bottom panel) data. In both plots, the data points represent grand averages across sets of items and trial repetitions, whereas the lines represent growth curve models, as detailed in the text

of the other four conditions were estimated relative to this condition. The statistical significance of the parameter estimates was assessed using the normal distribution.

As can be seen in Fig. 2, the segmental and cohort conditions differed from the baseline condition in both the experimental and simulation data, whereas the rhyme and tonal conditions did not. This is supported by the growth curve analyses; in the experimental data, the segmental and cohort conditions differed from the baseline condition in the cubic term of the model (segmental: *Estimate* = .091, *SE* = .049, *p* = .06; cohort: *Estimate* = .138, *SE* = .049, *p* < .01), whereas the rhyme and tonal conditions did not differ from the baseline condition in any term (all *p*s > .15). In the simulation data, the segmental and cohort conditions differed from the baseline condition in the intercept (segmental: *Estimate* = −.073, *SE* = .004, *p* < .001; cohort: *Estimate* = −.081, *SE* = .004, *p* < .001), linear (segmental: *Estimate* = −.114, *SE* = .014, *p* < .001; cohort: *Estimate* = −.089, *SE* = .014, *p* < .001), quadratic (segmental: *Estimate* = .421, *SE* = .022, *p* < .001; cohort: *Estimate* = .502, *SE* = .022, *p* < .001), cubic (segmental: *Estimate* = .158, *SE* = .017, *p* < .001; cohort: *Estimate* = .153, *SE* = .017, *p* < .001), and quartic (segmental: *Estimate* = −.177, *SE* = .010, *p* < .001; cohort: *Estimate* = −.252, *SE* = .010, *p* < .001) terms of the model, whereas the rhyme and tonal condition did not differ from the baseline condition in any term (all *p*s > .1).

On the basis of these results, we claim that the simulation and experimental data complement one another, as both showed evidence of competitive effects in the segmental and cohort conditions, and a lack of competitive effects in the rhyme and tonal conditions. As was argued in the original Malins and Joanisse (2010) article, the only difference between the segmental and cohort conditions was that segmental competitors differed from targets in tonal information, whereas cohort competitors differed from targets in vowels. For this reason, we argue that the model was able to capture within-syllable competitive effects arising due to either phonemic or tonal differences, echoing results from previous studies (Malins et al., 2014; Malins & Joanisse, 2012a; Zhao et al., 2011). The ability of the model to capture this pattern of effects illustrates its viability as a working model of spoken word recognition in Mandarin.

Nevertheless, we acknowledge that the overall shapes of the curves between the simulation and experimental data are not equivalent, which explains why different terms in the growth curve models showed differences across conditions. We believe the source of this discrepancy can be attributed to a key difference between the behavioral experiment and the simulations. Importantly, in the simulations we used a restricted lexicon in which there was at most one segmental competitor (i.e., an item with the same segmental structure but a different tone) for each item. However, for the behavioral experiment itself, competition presumably happened between

**Table 4** Sets of items in each condition of the tone competition simulation

| Tonal Contrast | Set | Target | Competitor | Distractor | Distractor |
|---|---|---|---|---|---|
| 1–2 | 1 | *bao1* | *bao2* | *fen3* | *luo4* |
| | 2 | *fen1* | *fen2* | *bao3* | *mi4* |
| | 3 | *luo1* | *luo2* | *mi4* | *fen3* |
| | 4 | *mi1* | *mi2* | *luo4* | *bao3* |
| 2–1 | 5 | *bao2* | *bao1* | *fen4* | *luo3* |
| | 6 | *fen2* | *fen1* | *luo3* | *bao4* |
| | 7 | *luo2* | *luo1* | *mi3* | *fen4* |
| | 8 | *mi2* | *mi1* | *luo4* | *bao3* |
| 1–3 | 1 | *bao1* | *bao3* | *luo2* | *mi4* |
| | 2 | *fen1* | *fen3* | *luo2* | *bao4* |
| | 3 | *luo1* | *luo3* | *fen4* | *bao2* |
| | 4 | *mi1* | *mi3* | *fen4* | *luo2* |
| 3–1 | 5 | *bao3* | *bao1* | *fen2* | *luo4* |
| | 6 | *fen3* | *fen1* | *mi2* | *luo4* |
| | 7 | *luo3* | *luo1* | *mi4* | *fen2* |
| | 8 | *mi3* | *mi1* | *luo4* | *bao2* |
| 1–4 | 1 | *bao1* | *bao4* | *mi2* | *fen3* |
| | 2 | *fen1* | *fen4* | *mi2* | *luo3* |
| | 3 | *luo1* | *luo4* | *bao3* | *mi2* |
| | 4 | *mi1* | *mi4* | *bao3* | *fen2* |
| 4–1 | 5 | *bao4* | *bao1* | *fen3* | *luo2* |
| | 6 | *fen4* | *fen1* | *bao2* | *mi3* |
| | 7 | *luo4* | *luo1* | *mi2* | *fen3* |
| | 8 | *mi4* | *mi1* | *luo3* | *bao2* |
| 2–3 | 1 | *bao2* | *bao3* | *luo4* | *mi1* |
| | 2 | *fen2* | *fen3* | *mi1* | *luo4* |
| | 3 | *luo2* | *luo3* | *fen1* | *bao4* |
| | 4 | *mi2* | *mi3* | *fen4* | *luo1* |
| 3–2 | 5 | *bao3* | *bao2* | *luo1* | *mi4* |
| | 6 | *fen3* | *fen2* | *bao1* | *mi4* |
| | 7 | *luo3* | *luo2* | *fen4* | *bao1* |
| | 8 | *mi3* | *mi2* | *fen4* | *luo1* |
| 2–4 | 1 | *bao2* | *bao4* | *mi3* | *fen1* |
| | 2 | *fen2* | *fen4* | *bao1* | *mi3* |
| | 3 | *luo2* | *luo4* | *bao1* | *mi3* |
| | 4 | *mi2* | *mi4* | *bao3* | *fen1* |
| 4–2 | 5 | *bao4* | *bao2* | *luo3* | *mi1* |
| | 6 | *fen4* | *fen2* | *luo1* | *bao3* |
| | 7 | *luo4* | *luo2* | *fen1* | *bao3* |
| | 8 | *mi4* | *mi2* | *fen3* | *luo1* |
| 3–4 | 1 | *bao3* | *bao4* | *mi1* | *fen2* |
| | 2 | *fen3* | *fen4* | *luo1* | *bao2* |
| | 3 | *luo3* | *luo4* | *bao2* | *mi1* |
| | 4 | *mi3* | *mi4* | *bao2* | *fen1* |
| 4–3 | 5 | *bao4* | *bao3* | *mi2* | *fen1* |
| | 6 | *fen4* | *fen3* | *mi1* | *luo2* |
| | 7 | *luo4* | *luo3* | *bao1* | *mi2* |
| | 8 | *mi4* | *mi3* | *bao2* | *fen1* |

Each set was used as a four-item tuple in the simulations. Reciprocal tonal contrast pairs (e.g., 1–2 and 2–1) were averaged together to create the grand average curves shown in Fig. 3

all items in the subjects' mental lexicon, rather than just the local set of items used in the experiment. As a result, the behavioral study was subject to additional psycholinguistic effects that could have influenced listeners' expectations for hearing certain syllables.

First, the stimulus set used in Malins and Joanisse (2010) did not contain examples of all possible tonal contrasts. For example, the tone 2 versus tone 3 contrast (subsequently denoted as *tone 2–3*) was lacking in the Malins and Joanisse (2010) study, which could have affected listeners' expectations for hearing certain tones when presented with either tone 2 or tone 3 targets. Second, the Malins and Joanisse (2010) stimulus set contained certain tonotactic gaps. More specifically, some of the syllables in the set did not contain entries in the mental lexicon with certain tones (e.g., there is no entry in the lexicon for the syllable *qiu* with the fourth tone, since this syllable–tone combination does not exist in Mandarin; Duanmu, 2007). Similar to the lack of counterbalancing of tonal contrast pairs, these tonotactic gaps in the stimulus set could have biased listeners' tonal expectations in a manner that was not captured by the simulations.

To extend the viability of TRACE-T, we wished to test whether the model is capable of simulating these two effects.

To do this, we simulated two additional visual world paradigm experiments inspired by previously published studies (e.g., Wiener & Ito, 2014). In the first simulation, we looked at the resolution of tonal competition for different tonal contrast pairs. To perform this simulation, we took four syllables that contained entries in the lexicon for all four tones. These four syllables contained different types of initial consonants (one nasal, one glide, one stop, and one fricative), as well as different vowel clusters. We devised a stimulus set in which each syllable appeared as a target in each tone, along with its respective segmental competitor in each of the other three tones (Table 4). Distractor items were selected from amongst the other items in the set so that they contained neither phonemic nor tonal overlap with the targets. We then simulated a visual world paradigm experiment using the same set of parameters as we detailed earlier (with the exception of the stochasticity parameter, which was set to 0) and a restricted lexicon consisting of only the items listed in Table 4.

The grand averages for the simulation are shown in Fig. 3. As is apparent in the figure, this simulation gave rise to a delay for tone 2–3 pairs as compared to the other contrasts in terms of the trajectory of looks to target. This replicates published findings in the literature, where it has been shown that

## Looks to Target for the Tone Competition Simulation
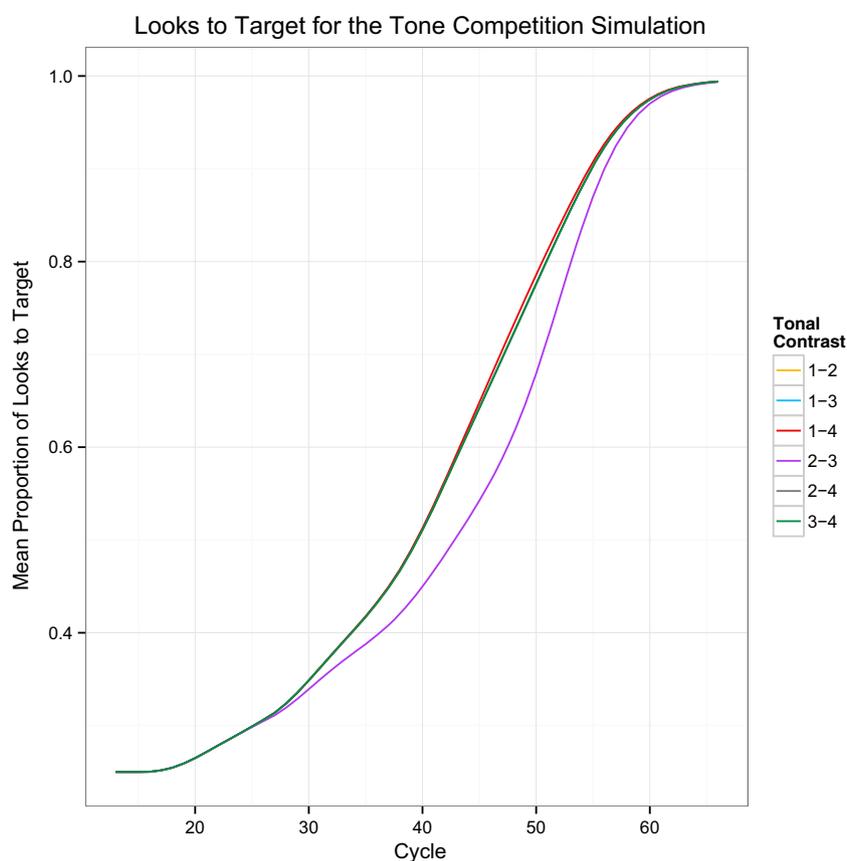
**Fig. 3** Mean proportions of looks to target for the tone competition simulation, as calculated by averaging together the results for each of the sets listed in Table 4. Note that the curves were generated by averaging together reciprocal tonal contrast pairs (e.g., "1–2" is the average of the sets with tone 1 as a target and tone 2 as a competitor and of the sets with tone 2 as a target and tone 1 as a competitor)

compared to other tonal contrast pairs, tones 2 and 3 compete with one another more strongly (Chandrasekaran, Krishnan, & Gandour, 2007), because they are discriminated on the basis of later acoustic cues during the unfolding of the syllable (Jongman, Wang, Moore, & Sereno, 2006; Shen, Deutsch, & Rayner, 2013; Whalen & Xu, 1992).

In the second simulation, we investigated the interaction between syllable frequency and tonal probability. *Tonal probability* refers to the likelihood that a syllable will be articulated in a certain tone in a tonal language. More specifically, it can be defined as the proportion of a syllable's total frequency (summed across all possible tones) divided by the frequency of the syllable articulated in a specific tone. For example, in Mandarin the tonal probability of tone 1 for the syllable *bao* is defined as the frequency of *bao1* divided by the summed frequency of *bao1*, *bao2*, *bao3*, and *bao4*. Recently, Wiener and Ito (2014) performed a visual world paradigm experiment in which high and low frequency target syllables were presented with segmental competitors. Importantly, the authors manipulated the tonal probability of the targets and competitors, such that the targets were high in probability and the competitors were low in probability, or vice versa. They found that tonal probability had an effect on the trajectory of looks to target only for low-frequency syllables.

To simulate this experimental effect, we devised a set of stimuli similar to those used in Wiener and Ito (2014) with some minor modifications. Using the *Modern Chinese Frequency Dictionary* (Beijing Language Institute, 1986), we selected syllables with either high or low logarithmic frequency (high = above 3.88, low = below 3.09; note that the mean logarithmic frequency of all items in the dictionary was 3.34), and then chose either high or low probability tones for each syllable (high = between 60 % and 90 %, low = between 5 % and 15 %). The full set of items is listed in Table 5. As can be noted from the table, the four conditions (syllable frequency high, tonal probability high, or FH–PH; syllable frequency high, tonal probability low, or FH–PL; and so on) were balanced in terms of tonal contrast pairs; furthermore, distractor items were selected from within the same set of stimuli in order to have a closed set. The simulations were performed as detailed above, with a few notable changes. To simulate frequency effects, we used the recommendations of Dahan, Magnuson, and Tanenhaus (2001) and changed the following parameters: Resting activation values of the lexical units were set to .06, phoneme to word level weights were set to .13, and the postactivation parameter was set to 15. To incorporate tonal probability into the model, we rescaled the frequencies using the following formula: $F_{new} = f - C/p$, where $F_{new}$ is an item's rescaled frequency, $f$ is the syllable frequency (summed across all four tones), $C$ is a constant (which we set to 100), and $p$ is the probability of a specific tone. Finally, stochasticity was again set to 0, and a restricted lexicon was used that contained only the items listed in Table 5.

**Table 5** Sets of items in each condition of the syllable frequency by tonal probability simulation

| Condition | Set | Target | Competitor | Distractor | Distractor |
|---|---|---|---|---|---|
| FH–PH | 1 | chu1 | chu2 | ren4 | ke3 |
| | 2 | jing1 | jing3 | qu4 | ren2 |
| | 3 | sheng1 | sheng4 | chu2 | hai3 |
| | 4 | tong2 | tong1 | hao3 | qu4 |
| | 5 | hai2 | hai3 | chu1 | ren4 |
| | 6 | ren2 | ren4 | xiao1 | zuo3 |
| | 7 | xiao3 | xiao1 | tong2 | ke4 |
| | 8 | hao3 | hao2 | sheng1 | qu4 |
| | 9 | ke3 | ke4 | jing1 | hai2 |
| | 10 | qu4 | qu1 | hai3 | ren2 |
| | 11 | da4 | da2 | jing3 | chu1 |
| | 12 | zuo4 | zuo3 | da2 | jing1 |
| FH–PL | 1 | chu2 | chu1 | ren4 | ke3 |
| | 2 | jing3 | jing1 | qu4 | ren2 |
| | 3 | sheng4 | sheng1 | chu2 | hai3 |
| | 4 | tong1 | tong2 | hao3 | qu4 |
| | 5 | hai3 | hai2 | chu1 | ren4 |
| | 6 | ren4 | ren2 | xiao1 | zuo3 |
| | 7 | xiao1 | xiao3 | tong2 | ke4 |
| | 8 | hao2 | hao3 | sheng1 | qu4 |
| | 9 | ke4 | ke3 | jing1 | hai2 |
| | 10 | qu1 | qu4 | hai3 | ren2 |
| | 11 | da2 | da4 | jing3 | chu1 |
| | 12 | zuo3 | zuo4 | da2 | jing1 |
| FL–PH | 1 | zhou1 | zhou2 | shai3 | qia4 |
| | 2 | shuang1 | shuang3 | zhou2 | bin4 |
| | 3 | bin1 | bin4 | shai3 | cao2 |
| | 4 | za2 | za1 | bin4 | peng3 |
| | 5 | peng2 | peng3 | qia4 | tie1 |
| | 6 | chen2 | chen4 | niao3 | za1 |
| | 7 | tie3 | tie1 | shai4 | cao2 |
| | 8 | cao3 | cao2 | chen4 | bin1 |
| | 9 | niao3 | niao4 | peng2 | zhou1 |
| | 10 | qia4 | qia1 | chen2 | peng3 |
| | 11 | kang4 | kang2 | zhou1 | tie3 |
| | 12 | shai4 | shai3 | tie1 | zhou2 |
| FL–PL | 1 | zhou2 | zhou1 | shai3 | qia4 |
| | 2 | shuang3 | shuang1 | zhou2 | bin4 |
| | 3 | bin4 | bin1 | shai3 | cao2 |
| | 4 | za1 | za2 | bin4 | peng3 |
| | 5 | peng3 | peng2 | qia4 | tie1 |
| | 6 | chen4 | chen2 | niao3 | za1 |
| | 7 | tie1 | tie3 | shai4 | cao2 |
| | 8 | cao2 | cao3 | chen4 | bin1 |
| | 9 | niao4 | niao3 | peng2 | zhou1 |
| | 10 | qia1 | qia4 | chen2 | peng3 |
| | 11 | kang2 | kang4 | zhou1 | tie3 |
| | 12 | shai3 | shai4 | tie1 | zhou2 |

Each set was used as a four-item tuple in the simulations. FH = high syllable frequency, FL = low syllable frequency, PH = high tonal probability, PL = low tonal probability
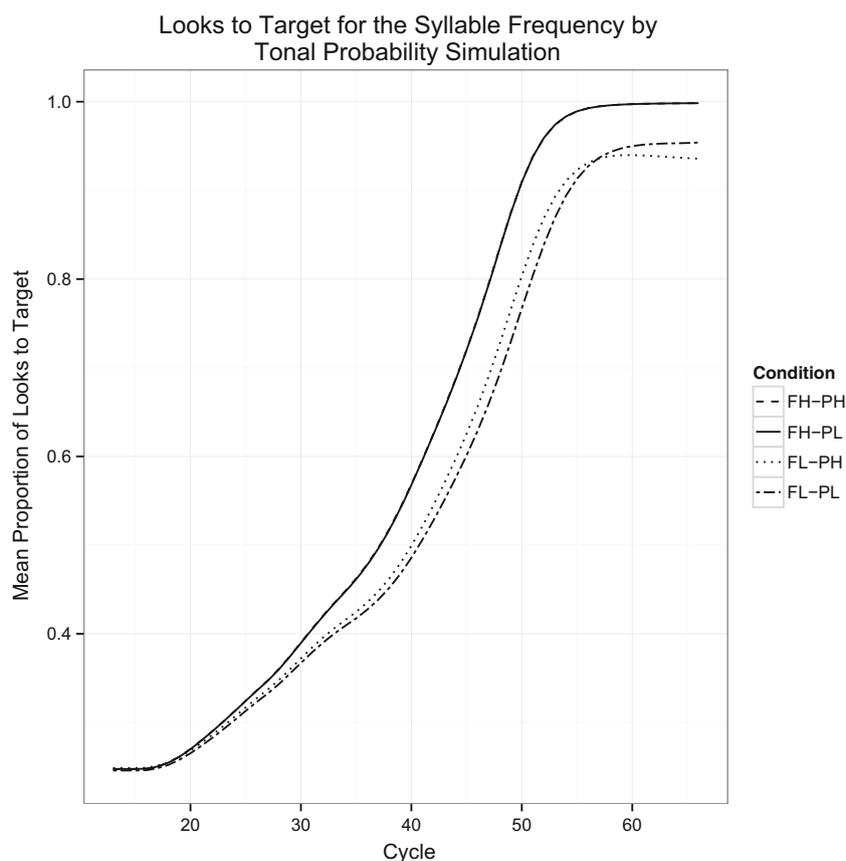
Fig. 4 Mean proportions of looks to target for the syllable frequency by tonal probability simulation, as calculated by averaging together the results for each of the sets listed in Table 5. FH = high syllable frequency, FL = low syllable frequency, PH = high tonal probability, PL = low tonal probability

The grand averages from the simulations are shown in Fig. 4. As can be discerned from the figure, we observed an interaction between syllable frequency and tonal probability: The trajectory of looks to target in the FH–PH condition is indistinguishable from the trajectory for the FH–PL condition, yet the FL–PH condition shows a slight advantage as compared to the FL–PL condition. This interaction is in the same direction as that reported by Wiener and Ito (2014), suggesting that TRACE-T is sensitive to this pattern of effects.

Together, these additional simulations provide further evidence that TRACE-T is capable of simulating documented psycholinguistic effects in Mandarin spoken word recognition. Namely, TRACE-T was able to simulate increased competitive effects for tone 2–3 pairs as well as the interaction between syllable frequency and tonal probability. These findings bolster our confidence that the model simulates Mandarin spoken word recognition in a reasonable fashion, and that differences between the Malins and Joanisse (2010) experimental data and the simulation results can be attributed to additional psycholinguistic effects beyond those captured in the simulation, rather than to deficiencies in the model architecture itself. Furthermore, TRACE-T has given us valuable insights by showing what the Malins and Joanisse (2010) eyetracking results might have

looked like in the absence of these task-specific effects. This serves as a powerful illustration of the ability of computational models such as TRACE-T to illuminate patterns in experimental data and to serve as sources of information that can be considered complementary to experimental results.

## Conclusions

As is detailed in this report, we have developed a methodology that allows researchers to simulate spoken word recognition in Mandarin Chinese. We took the existing jTRACE platform, which is available and easy to use (Strauss et al., 2007), and modified it such that it codes Mandarin phonemes and tones in a way that is reasonable on the basis of previous findings (e.g., Brown-Schmidt & Canseco-Gonzalez, 2004; Malins et al., 2014; Malins & Joanisse, 2010, 2012a; Schirmer et al., 2005; Zhao et al., 2011). We then validated the model by replicating published eyetracking data from human subjects. Now that we have proposed an initial working model—which we call TRACE-T, in reflection of its ability to encode lexical tone—future research should help refine this model, which can then in turn be used to generate new

hypotheses and motivate future studies regarding Mandarin Chinese speech perception. Furthermore, we hope this model will help inspire computational models for other tonal languages, such as Cantonese and Thai. Together, these kinds of endeavors can help push forward psycholinguistic research by allowing speech perception theories to accommodate data from an increasingly greater proportion of the world's languages.

# References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439. doi:10.1006/jmla.1997.2558

Beijing Language Institute. (1986). *Modern Chinese frequency dictionary [in Chinese]*. Beijing: Beijing Language Institute.

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research, 33,* 103–135.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007a). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25,* 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128,* 148–156. doi:10.1016/j.brainres.2006.10.064

Chen, T. Y., & Tucker, B. V. (2013). Sonorant onset pitch as a perceptual cue of lexical tones in Mandarin. *Phonetica, 70,* 207–239.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York: Oxford University Press.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12,* 235–261.

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

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 209–217). Cambridge: Cambridge University Press.

Li, P., Tan, L. H., Bates, E., & Tzeng, O. J. L. (Eds.). (2006). *The handbook of East Asian psycholinguistics: Vol. 1. Chinese*. New York: Cambridge University Press.

Malins, J. G., Gao, D., Tao, R., Booth, J. R., Shu, H., Joanisse, M. F., Liu, L., & Desroches, A. S. (2014). Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese. *Brain and Language, 138,* 38–50. doi:10.1016/j.bandl.2014.09.002

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62,* 407–420. doi:10.1016/j.jml.2010.02.004

Malins, J. G., & Joanisse, M. F. (2012a). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50,* 2032–2043. doi:10.1016/j.neuropsychologia.2012.05.002

Malins, J. G., & Joanisse, M. (2012b). *Towards a model of tonal processing during Mandarin spoken word recognition*. Paper presented at the Third International Symposium on Tonal Aspects of Languages, Nanjing, China.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101,* 653–675. doi:10.1037/0033-295X.101.4.653

Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from TRACE simulations. *Journal of Memory and Language, 71,* 89–123. doi:10.1016/j.jml.2013.09.009

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86. doi:10.1016/0010-0285(86)90015-0

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology, 60,* 1–39. doi:10.1016/j.cogpsych.2009.06.003

Mirman, D. (2014). *Growth curve analysis and visualization using R*. New York: CRC Press.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language, 117,* 53–68. doi:10.1016/j.bandl.2011.01.004

Protopapas, A. (1999). Connectionist modeling of speech perception. *Psychological Bulletin, 125,* 410–436. doi:10.1037/0033-2909.125.4.410

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17,* 1–12. doi:10.1162/0898929052880057

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin tones 2 and 3: Evidence from eye movements. *Journal of the Acoustical Society of America, 133,* 3016–3029.

Shuai, L., Gong, T., Ho, J. P.-K., & Wang, W. S.-Y. (2016). Hemispheric lateralization of perceiving Cantonese contour and level tones: An ERP study. In W. Gu (Ed.), *Studies on tonal aspect of languages (Journal of Chinese Linguistics Monograph Series, No. 25)*. Hong Kong: Journal of Chinese Linguistics, in press.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39,* 19–30. doi:10.3758/BF03192840

Tong, X., McBride, C., & Burnham, D. (2014). Cues for lexical tone perception in children: Acoustic correlates and phonetic context effects. *Journal of Speech, Language, and Hearing Research, 57,* 1589–1605. doi:10.1044/2014_JSLHR-S-13-0145

Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33,* 93–105.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25–47.

Wiener, S., & Ito, K. (2014). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience, 30,* 1048–1060. doi:10.1080/23273798.2014.946934

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language & Cognitive Processes, 14,* 609–630.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia, 49,* 1761–1770. doi:10.1016/j. neuropsychologia.2011.02.054

Zhao, X., & Li, P. (2009). An online database of phonological representations for Mandarin Chinese. *Behavior Research Methods, 41,* 575–583. doi:10.3758/BRM.41.2.575

Zhou, X., & Shu, H. (2011). Neurocognitive processing of the Chinese language. *Brain and Language, 119,* 59. doi:10.1016/j.bandl.2011. 06.003

Zhou, N., Zhang, W., Lee, C. Y., & Xu, L. (2008). Lexical tone recognition with an artificial neural network. *Ear and Hearing, 29,* 326–335.