# L2 PHONOLOGICAL CONTRASTS ARE NOT NECESSARILY LOST: EVIDENCE FROM FRENCH LISTENERS DETECTING MANDARIN TONES

Volker Dellwo, Sandra Schwab, Rushen Shi

University of Zurich, University of Bern, University of Quebec at Montreal
volker.dellwo@uzh.ch, sandra.schwab@unibe.ch, shi.rushen@uqam.ca

## ABSTRACT

Adult listeners are typically poor in perceiving non-native phonemic contrasts that are absent in their native language. French listeners, for example, whose native language lacks contrastive word stress, have a documented poor contrastive stress distinction ability. Here, we tested the performance of French listeners to detect tone contrasts, another phonological contrast that is not present in French. Forty-nine French listeners detected Mandarin level- and contour-tone contrasts in an odd-one-out task in either mono- or bisyllabic words. Performance was high and significantly above chance for all contrasts but contrasts in monosyllabic words were significantly better detected than in bisyllabic words. We conclude that the ability to detect non-native phonological contrasts depends highly on the nature of the contrast (here: stress-contrast vs. tone contrasts), a finding that might be important in the light of language evolution.

**Keywords**: L2 contrast perception, tone perception, stress perception, language evolution

## 1. INTRODUCTION

Early in development, infants possess an acute ability to perceive both native and non-native contrasts. The sensitivity to non-native contrasts, however, gradually declines towards the end of the first year of life while the perception of native contrasts continues ([1], [2]) or further strengthens ([3]). In French, for example, there is extensive evidence that adult listeners have very poor ability to distinguish contrastive stess in languages like Spanish or English, a phenomenon commonly known as *stressdeafness* ([4], [5]). The most likely reason given is that French phonology does not have a contrastive stress system, hence sensitive of listeners to such contrast is lost at some point during language acquisition.

Here, we tested how good French listeners are at distinguishing lexical tones, another phonological contrast that does not exist in French. In infants, contrastive tone detection performance in listeners from non-tonal languages varies. English- and French-learning infants discriminate rising versus falling contrast and the rising versus low contrast in Thai at 4 months of age, but the ability declines by 9 months ([6], [7]). Similarly, English-learning infants discriminate Cantonese high-rising and mid-level tones at 4 months of age but declined in this ability by 9 months ([8]). However, decline was not always observed. For example, French-learning infants aged 4 to 11 months showed no decline in discriminating the high-level versus falling contrast in Mandarin, while they showed a tendency of decline for another acoustically more similar tonal contrast, rising versus low-dipping ([9]). Similarly, Dutch-learning infants showed no decline in their discrimination of the Mandarin high-level versus falling tones during the first year of life, although a decline was observed when the acoustical difference of this contrast was artificially reduced ([10]). On the other hand, [11] showed that English-learning infants improved in their discrimination of a tonal contrast (high-level versus low-dipping) in Mandarin from 6 to 12 months of age, suggesting the possible contribution of growing general auditory processing capacities independent of the exposure to a pertinent phonemic system.

These results suggest that the perception of lexical tones by non-tonal language learners is less following the general decline of non-native contrasts replicated in typical non-native consonant and vowel perception tasks. Thus, it is unclear how the lack of lexical tone contrasts in a listener's native language affects its recognition ability. [12] tested 11-to-12-month-old French-learning infants on two types of tonal contrasts in Mandarin, level tone contrasts (high versus low, i.e., tone 1 vs tone 3; henceforth: T1, T3) and contour tone contrasts (rise versus fall, tone 2 vs tone 4; T2, T4). When tone contrasting syllables were the first syllable in bisyllabic pseudowords (second syllable was always T1), infants showed strong discrimination performance for both level and contour contrasts. However, when only the first syllable was presented to infants, performance dropped for contour tones and was at chance for level tones. The superior discrimination of both types of tonal contrasts when the bisyllabic forms were presented suggests that tonal differences between the first and the second syllable possibly offered salient acoustic cues. Importantly, the lack of lexical tones in infants' native language did not prevent them from perceiving tonal contrasts, at least under some conditions.

Here, we tested French native adults on the same stimuli that were used for the infants in [12] to understand whether the continued lack of experience with tone contrasts ultimately leads to a poor ability to perceive either level-tone (T1, T3) or contour-tone contrasts (T2, T4). If non-native language phonological contrasts decay with age, we expect French speakers to show poor discrimination performance for Mandarin tonal contrasts, for example similar to performances found for stress contrasts [5]. Additionally, we wanted to know (a) whether French adults – if able to detect Mandarin tones – rely on the second syllable or whether the tone in the first syllable is sufficient and (b) whether discrimination ability varies between contour and level contrasts. Given the larger degree of acoustic dynamic changes in contour contrasts, performance should be better compared to level contrasts that are mainly distinguished by tonal height differences.

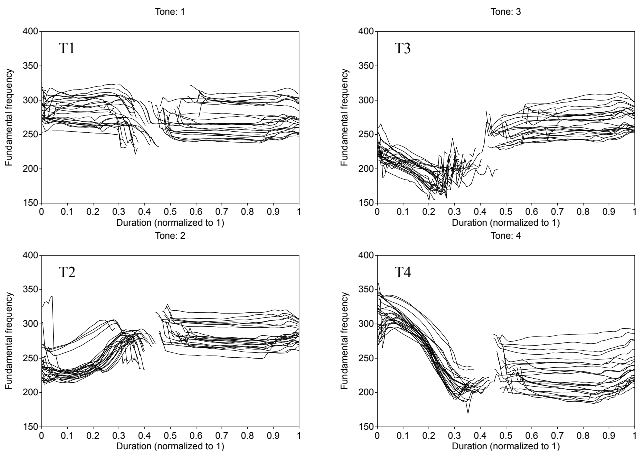


**Figure 1**: Pitch contours of all stimuli (with repetitions; N=96) separated by tones with level-tone examples (T1, T3; top) and contour tone examples (T2, T4; bottom). Duration has been normalized between 0 (onset of tone in first syllable) and 1 (offset of tone in last syllable).

## 2. METHOD

### 2.1. Participants

Forty-nine native listener of French from Switzerland and France (age range: 18 to 40, mean age = 27.27 years). No participant had knowledge of a tone language.

### 2.2. Material

***Stimuli:*** Six Mandarin Chinese bisyllabic pseudowords (*fadi, quigu, pengbu, wangpi, xianda, yuanta*) were recorded from a female native speaker of Mandarin Chinese. Each word was produced with T1 (high tone), T2 (rising tone), T3 (low tone) and T4 (falling tone) on the first syllable and with T1 on the second syllable which was always /di:/ (see example in Fig. 1). Four repetitions of each of the 24 pseudowords were produced by the native speaker resulting in a total of 96 stimuli (6 words * 4 tones * 4 repetitions).

***Trials:*** Trials of three stimuli were created in which one stimulus (odd) differed from the other two stimuli (even) regarding its tone. While the tone on the even stimuli was from the same category, the two stimuli were two randomly chosen realisations of the four repetitions. Combinations were either created with level- or contour tone contrasts, resulting in four combinations (Table 1). Two variants from different repetitions were created for the four tone combinations for each pseudoword resulting in 48 trials (4 tone combinations * 6 pseudowords * 2 variants). The three stimuli in a trial were separated by 500 ms of silence.

***Trial sets:*** A second trial set was created that was identical to the first set with the exception that the second syllable of the bisyllabic pseudoword stimuli was deleted (from the onset of the stop gap of the second syllable /di:/). In total, the experimental material thus consisted of two trial sets, one containing the bisyllabic words (henceforth: *All*) and the other containing only the first syllable (henceforth: *1st*).

| Contrasts | Odd | Even |
|---|---|---|
| Level (T1-T3) | 1 | 3 |
| Level (T1-T3) | 3 | 1 |
| Contour (T2-T4) | 2 | 4 |
| Contour (T2-T4) | 4 | 2 |

**Table 1**: Combinations of tones in experimental trials with two level tone combinations (rows 1, 2) and two contour tone combinations (rows 3, 4).

### 2.3. Procedure

Participants were randomly split in two groups, one group (N=30) listening to the *All* trial set the other (N=19) to *1st* trial set. In an odd-one-out task, participants were presented each of the 48 trials of the respective trial set via high-quality headphones connected to a computer sound card in a quiet classroom at the University of Geneva. After each trial presentation, listeners task was to detect the position of the odd stimulus (the stimulus varyin in tone from the other two) in each trial and indicate the sequential position in the trial (1st, 2nd or 3rd) on corresponding buttons on the screen. Combinations of stimuli within each trial and the sequence of trials in each trial set were fully randomised for each individual listener. Experimentas were run in Praat using scripts (www.praat.org) available from the first author.

### 2.4. Data analysis

Statistical analyses were carried out using R (version 3.4.0, [13]; R Development Core Team, 2016; lmerTest R package, [14]). A mixed-effects logistic regression model was run on the correct (odd detected)/incorrect (odd not deteceted) responses ([15]). The fixed part of the model was comprised of trial set (*All*, *1st*), tone contrast (level, contour), and their interaction. Both level contrasts and contour contrasts from Table 1 were respectively grouped together so that there were only two types of contrasts: level tone and contour tone contrasts. The random part of the model included random intercept for participants, and random slopes allowing for the effect of 'tone contrast' to differ across participants. The significance of the main effects and interaction was assessed with likelihood ratio tests that compared the model with the main effect or interaction to a model without it. For descriptive reasons, Fig. 2 presents percent correct, however, the analysis was performed on correct/incorrect responses.

In addition, participants' responses to each tonal contrast, i.e., T1 vs T3 (Level) and T2 vs T4 (Contour), in the *All* and *1st* contexts were compared with chance in exact binomial tests, allowing us to assess if the two types of tonal contrasts were perceived in these contexts by adults.

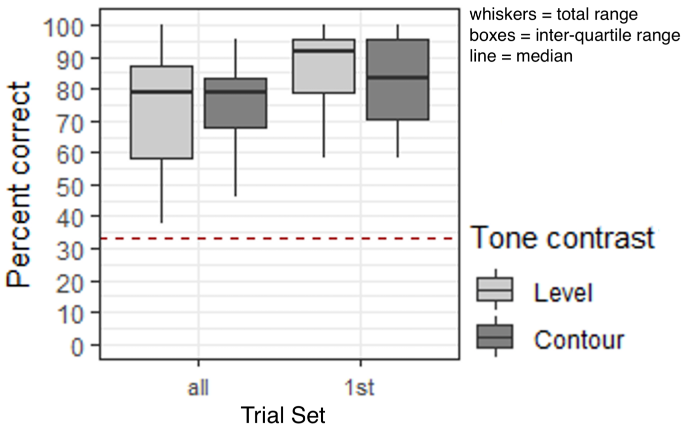


**Figure 2**: Percent correct as a function of *Trial Set* (All, first syllable) and 'tone contrast' (level, contour). The dashed line represents chance level (33%).

## 3. RESULTS

Fig. 2 presents the percent correct as a function of stimulus set and tone contrast. Performance for each tonal contrast in the two trial sets was overall high and each performance was significantly above the 33.3% chance level (All Level = 74%, exact binomial p [one-tailed] < 0.001; All Contour = 77%, exact binomial p [one-tailed] < 0.001; Level = 86%, exact binomial p [one-tailed] < 0.001; Contour = 82%, exact binomial p [one-tailed] < 0.001).

The interaction between 'trial set' and 'tone contrast' was close to significance (Tab. 2). For this reason we carried out simple effect post-hoc analyses (with Tukey correction) rather then interpretation of main effects. Simple effect analysis revealed that the difference between the two types of tonal contrasts (i.e., Level, Contour) was not significant in *All* (Level = 74%; Contour = 77%; β = 0.172, SE = 0.127, z = 1.351, p = .177) nor in *1st* (Level = 86%; Contour = 82%; β = 0.327, SE = 0.189, z = -1.728 , p = .084). Simple effect analyses also showed that the performance for contour tones did not significantly differ across the two stimulus sets (All-1st: β = -0.433, SE = 0.292, z = -1.486, p = .137). On the contrary, for level tones, the performance was better with the extracted first syllable (86%) than with the entire bisyllabic pseudoform (74%) (β = 0. 933, SE = 0.297, z = 3. 138, p = .002)

| *Predictors* | *Log-Odds* | *CI* | *p* | *Likelihood ratio test* |
|---|---|---|---|---|
| (Intercept) | 1.243 | 0.840 – 1.646 | **<0.001** | |
| **wordPart (ref = all)** | | | | $\chi^2(1) = 7.288$, p = .007 |
| 1st | 0.952 | 0.280 – 1.623 | **0.005** | |
| **toneContrast (ref = level)** | | | | $\chi^2(1) = 0.176$, p = .675 |
| Contour | 0.070 | -0.251 – 0.391 | 0.668 | |
| **wordPart x toneContrast** | | | | $\chi^2(1) = 3.411$, p = .065 |
| 1st:Contour | -0.531 | -1.076 – 0.015 | 0.057 | |

**Table 2**: Summary of mixed-effects logistic regression model with likelihood ratio tests. Reference is 'incorrect' for the dependent variable, 'All' for wordPart and 'Level' for toneContrast.

## 3. DISCUSSION

We obtained a high detection rate of Mandarin tone-contrasts by French listeners in bisyllabic pseudowords (~75%) and significantly stronger performance when only the tone bearing syllable was presented (~84%). This suggests that the lack of exposure to tonal contrasts in their native language did not considerably impede French listener's detection performance of Mandarin Chinese tone contrasts. Thus non-native tone contrasts do not necessarily decay with age as they do with stress-contrasts in French ([4], [5]).

How much better are French listeners in detecting tone contrasts compared to stress contrasts? In a contrastive word-stress detection task using an odd-one-out design, [5] tested French listeners – amongst others – on their ability to detect stress differences in single talker and double talker condition. [5] found that using one talker and one intonation – identical to the present experiment – yielded ~63% correct for stress-contrast detection. This result is drastically lower than any of the conditions tested in the present study and is strong evidence that for French listeners the ability to detect tone is generally better maintained compared to the ability to detect stress-contrasts.

These findings are interesting in the light of the evolution of phonological contrasts in language systems, since they suggest that some contrasts are acoustically more prone to be stable throughout the lifetime of a listener while others seem to be more subtle and more likely to be lost. Stress-contrasts, for example, rely on a detection of a combination of cues like duration, pitch and loudness, while fundamental frequence is dominant in tone detection (at least for the stimuli in our experiment; not necessarily for tone languages in general). It seems possible that contrasts which rely on an auditorily less complex interplay of acoustic parameters might be more stable throughout the lifetime. This might be a reason for the vast success of tone systems around the world, in particular for the adoption of tones across typologically non-related languages during tono-genisis in Southeast Asia [16], [17].

How robust are the findings of the present study? In [5] French listeners were tested under additional conditions: (a) there was a single and double speaker design and (b) intonation variability was controlled in having either one statement intonation or varying between a statement and a question intonation contour. Interestingly, [5] found that performance dropped significantly when a second talker was introduced and was at chance level with two intonation contours. We thus expect that performance in tone perception might significantly drop when the task gains in difficulty, e.g. by increasing the number of speakers, styles, phrase positions where the tone contrasts occur, etc. In fact, it seems obvious from our results that the use of a single speaker might have led to a disproportional advantage for adults. As can be seen in Fig. 1, f0 is systematically higher than 250 Hz in T1 and lower then 250 Hz in T3 for this particular speaker. This means that in an odd-one-out task, a tone higher than 250 Hz is easy to contrast against two others that are lower than 250 Hz. It would be interesting to test French listener performance when speakers in the stimulus trials vary (e.g. male and female with drastically different average f0). A native control group would be necessary for such a task as it is unclear whether native Mandarin listeners would overcome such strong difference in level tones in single syllable presentation.

What does the finding reveal about the tone discrimination ability in adults versus infants? The test procedures in [12] were very different from the odd-one-out design in the present experiment which must necessarily be the case. The infants in [12] were tested with a preferential head turning procedure, a method by which the infant is first habituated for 30 seconds with one target tone (in various bisyllables or monosyllables). Subsequently, by presenting a novel contrast versus new exemplars of the familiarized tone in separate trials, it can be tested whether the infant recognizes the earlier familiarized tone as distinct from the novel tone by turning their head to the respective sound source and yielding systematic differences in looking durations. Here, adult listeners were tested with an odd-one-out design in which they identified the odd stimulus that carried a different tone compared to two even stimuli in a three stimuli trial. This might also explain why adults were better with single syllable task compared to the double while for infants this was the other way round. Given the lack of comparative information in the single syllable condition, it seems somehow counter intuitive that adults revealed a higher performance. Again, this might very likely be related to the single speaker design in which adults might have gained a disproportional advantage in the odd-one-out task compared to the head-turning task in infants. For a preferential head-turning task, the nature of a simple tone height difference (i.e., citation tones in monosyllables) might not be interesting enough for the infant to change their attention levels. In contrast, the fact that infants responded strongly to contour tone differneces may reveal that a dynamic difference between the direction of F0 change might increase attention sufficiently. For adults, this might mean that not their ability to detect abstract Mandarin tone categories led to their high performance in the single syllable condition but the fact that the tone differences in the trials were auditorily well distinguishable. We take this as further evidence that general, non-language or speech specific factors contributed to the performance because adding more speech detail led to confusion and a drop in performance. This means that the auditory clarity between the tone contours in a single syllable outweighed additional language specific information. In native listeners, we would expect this effect to be reversed.

In summary, further experiments will be necessary by using multiple speakers, stimulus variability, listener groups with varying L2 background and control groups. This will be interesting to understand the maintenance of non-L1 phonolgical contrasts. It will have high impact on the understanding of the development of contrast perception throughout the human lifetime and the evolution of phonological systems of languages. It will also be interesting to compare tone perception mechnisms across speech and non-speech sounds [18].

## 5. REFERENCES


[1] J. F. Werker and R. C. Tees, “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,” *Infant Behav Dev*, vol. 7, no. 1, pp. 49–63, 1984, doi: 10.1016/S0163-6383(84)80022-3.

[2] L. Polka and J. F. Werker, “Developmental Changes in Perception of Nonnative Vowel Contrasts,” *J Exp Psychol Hum Percept Perform*, vol. 20, no. 2, pp. 421–435, 1994, doi: 10.1037/0096-1523.20.2.421.

[3] P. K. Kuhl, E. Stevens, A. Hayashi, T. Deguchi, S. Kiritani, and P. Iverson, “Infants show a facilitation effect for native language phonetic perception between 6 and 12 months,” *Dev Sci*, vol. 9, no. 2, pp. F13–F21, Mar. 2006, doi: 10.1111/J.1467-7687.2006.00468.X.

[4] E. Dupoux, C. Pallier, N. Sebastian, and J. Mehler, “A Destressing ‘“Deafness”’ in French?,” 1997.

[5] S. Schwab and V. Dellwo, “Intonation and talker variability in the discrimination of Spanish lexical stress contrasts by Spanish, German and French listeners,” *J Acoust Soc Am*, vol. 142, no. 4, pp. 2419–2429, 2017, doi: 10.1121/1.5008849.

[6] K. Mattock and D. Burnham, “Chinese and English infants’ tone perception: Evidence for perceptual reorganization,” *Infancy*, vol. 10, no. 3, pp. 241–265, 2006, doi: 10.1207/S15327078IN1003_3.

[7] K. Mattock, M. Molnar, L. Polka, and D. Burnham, “The developmental course of lexical tone perception in the first year of life,” *Cognition*, vol. 106, no. 3, pp. 1367–1381, Mar. 2008, doi: 10.1016/J.COGNITION.2007.07.002.

[8] H. H. Yeung, K. H. Chen, and J. F. Werker, “When does native language input affect phonetic perception? The precocious case of lexical tone,” *J Mem Lang*, vol. 68, no. 2, pp. 123–139, Feb. 2013, doi: 10.1016/J.JML.2012.09.004.

[9] R. Shi, E. Santos, J. Gao, and A. Li, “Perception of Similar and Dissimilar Lexical Tones by Non-Tone-Learning Infants,” *Infancy*, vol. 22, no. 6, pp. 790–800, Nov. 2017, doi: 10.1111/INFA.12191.

[10] L. Liu and R. Kager, “Perception of tones by infants learning a non-tone language,” *Cognition*, vol. 133, no. 2, pp. 385–394, Nov. 2014, doi: 10.1016/J.COGNITION.2014.06.004.

[11] F. M. Tsao, “Perceptual improvement of lexical tones in infants: Effects of tone language experience,” *Front Psychol*, vol. 8, no. APR, p. 558, Apr. 2017, doi: 10.3389/FPSYG.2017.00558/BIBTEX.

[12] E. Santos and R. Shi, “Perception of level tones and contour tones by French-learning infants. ,” in *14th International Congress for the Study of Child Language (IASCL)*, Lyon, 2017.

[13] “R: a language and environment for statistical computing.” https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing (accessed Dec. 17, 2022).

[14] “Tests in Linear Mixed Effects Models [R package lmerTest version 3.1-3],” Oct. 2020, Accessed: Dec. 18, 2022. [Online]. Available: https://CRAN.R-project.org/package=lmerTest

[15] R. H. Baayen, D. J. Davidson, and D. M. Bates, “Mixed-effects modeling with crossed random effects for subjects and items,” *J Mem Lang*, vol. 59, no. 4, pp. 390–412, Nov. 2008, doi: 10.1016/J.JML.2007.12.005.

[16] J. A. Matisoff, “Tonogenesis in Southeast Asia,” *Consonant Types and Tone. Southern California Occasional Papers in Linguistics*, vol. 1. pp. 73–95, 1973.

[17] A. S. Abramson, “The plausibility of phonetic explanations of tonogenesis,” in *From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji’s 95th birthday*, G. F. Fant Hiroya; Cao, Jianfen; Xu, Yi, Ed., Beijing: Foreign Language Teaching and Research Press, 2004, pp. 17–29.

[18] T. Bent, A. R. Bradlow, and B. A. Wright, “The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds.,” *J Exp Psychol Hum Percept Perform*, vol. 32, no. 1, pp. 97–103, 2006, doi: 10.1037/0096-1523.32.1.97.