frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 12 July 2022
doi: 10.3389/fpsyg.2022.902569

The Sequence Recall Task and Lexicality of Tone: Exploring Tone "Deafness"

*Carlos Gussenhoven[1,2], Yu-An Lu[2], Sang-Im Lee-Kim[2], Chunhui Liu[3], Hamed Rahmani[1], Tomas Riad[4] and Hatice Zora[5]**

[1]Centre for Language Studies, Radboud University, Nijmegen, Netherlands, [2]Department of Foreign Languages and Literatures, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, [3]College of Literature and Journalism, Sichuan University, Chengdu, China, [4]Department of Swedish Language and Multilingualism, Stockholm University, Stockholm, Sweden, [5]Department of Neurobiology of Language, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Many perception and processing effects of the lexical status of tone have been found in behavioral, psycholinguistic, and neuroscientific research, often pitting varieties of tonal Chinese against non-tonal Germanic languages. While the linguistic and cognitive evidence for lexical tone is therefore beyond dispute, the word prosodic systems of many languages continue to escape the categorizations of typologists. One controversy concerns the existence of a typological class of "pitch accent languages," another the underlying phonological nature of surface tone contrasts, which in some cases have been claimed to be metrical rather than tonal. We address the question whether the Sequence Recall Task (SRT), which has been shown to discriminate between languages with and without word stress, can distinguish languages with and without lexical tone. Using participants from non-tonal Indonesian, semi-tonal Swedish, and two varieties of tonal Mandarin, we ran SRTs with monosyllabic tonal contrasts to test the hypothesis that high performance in a tonal SRT indicates the lexical status of tone. An additional question concerned the extent to which accuracy scores depended on phonological and phonetic properties of a language's tone system, like its complexity, the existence of an experimental contrast in a language's phonology, and the phonetic salience of a contrast. The results suggest that a tonal SRT is not likely to discriminate between tonal and non-tonal languages within a typologically varied group, because of the effects of specific properties of their tone systems. Future research should therefore address the first hypothesis with participants from otherwise similar tonal and non-tonal varieties of the same language, where results from a tonal SRT may make a useful contribution to the typological debate on word prosody.

Keywords: word prosody, lexicon-based memory, tone contrast salience, tone language, semi-tonal language, sequence recall task

OPEN ACCESS

Edited by:
Juhani Järvikivi,
University of Alberta, Canada

Reviewed by:
Ting Zou,
Beijing Foreign Studies University, China
Ratree Wayland,
University of Florida, United States

****Correspondence:***
Hatice Zora
hatice.zora@mpi.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *23 March 2022*
Accepted: *16 May 2022*
Published: *12 July 2022*

Citation:
Gussenhoven C, Lu Y-A, Lee-Kim S-I, Liu C, Rahmani H, Riad T and Zora H (2022) The Sequence Recall Task and Lexicality of Tone: Exploring Tone "Deafness". Front. Psychol. 13:902569. doi: 10.3389/fpsyg.2022.902569

INTRODUCTION

Lexical tone has been investigated in a large body of perception research and is a prominent traditional typological concept in phonology, perhaps more so than word stress, which until recently was often treated as a universal (*cf.* van Heuven and Turk, 2020). Tones can form a great variety of subsystems in the phonologies of languages. There can be few or many of

them and contrasts will vary in salience. Functionally, they could share the phonological specification of morphemes with vowels and consonants ("lexical tones") or be their sole exponents ("grammatical tones," Hyman, 2011, 2016). While the linguistic and cognitive evidence for lexical tone is beyond dispute, as indicated by the results of dichotic listening, categorical perception, ABX designs, and brain response registrations (Lau et al., 2020), the word prosodic systems of many languages continue to escape the categorizations of typologists, with frequent debates about the categorization of tone languages (Hyman, 2006; Kehrein et al., 2017; Steien and Yakpo, 2020; Gooden, 2022). The present paper aims to contribute to the understanding of the lexical status of tone by comparing non-tonal, semi-tonal, and tonal languages in a Sequence Recall Task (SRT). It was developed by Emmanuel Dupoux and colleagues as a diagnostic for the presence of word stress in a language (Dupoux et al., 2001). It followed their earlier speculations on why French listeners underperformed in an ABX task relative to Spanish listeners, where A and B were trisyllabic non-words differing in the location of stress (Dupoux et al., 1997). An SRT trial presents participants with a sequence of some 4 to 6 disyllabic non-words which have a prominence on either one or another of its syllables, as in the disyllabic non-word sequence *númi – numí – númi – númi*. Participants are asked to reproduce the order of the two non-words on a keyboard (in this case 1–2–1–1) after hearing a distracting sound immediately after the sequence, intended to prevent them from relying on their acoustic memory (cf. Baddeley, 2010). Speakers of Spanish, a language with contrastive word stress, outperformed speakers of French on this task, which language has phrasal stress (Dupoux et al., 2001). The effect survives language contact as in L2 learning (Dupoux et al., 2008).

Explanations of the inability of French listeners to perform the task as effectively as Spanish listeners first addressed the exposure to meaningful word prosody during language acquisition, but later shifted to the resulting abstract lexical representation of stress (Peperkamp, 2004; Dupoux et al., 2008). Providing support for this interpretation, Rahmani et al. (2015) showed that the presence of syllabic prominence in lexical representations, whether from tone or stress, explained the results of an experiment with five language groups, Dutch, Japanese, French, Indonesian, and Persian. As hypothesized, Dutch and Japanese participants outperformed the participants in the other three language groups, who for that reason are "stress-deaf" (the term is due to Dupoux et al., 1997). The explanation the authors give is that Dutch and Japanese participants could engage their lexicon-based memory on the basis of the contrastive location of a syllabic prominence in words, stress in Dutch and a HL melody in Japanese. The interpretation of stress as tone by the Japanese listeners was also evident in Qin et al. (2017), in which Standard Mandarin, Taiwan Mandarin, and English participants achieved comparable SRT performance on disyllabic English stress pairs. None of the other three languages in Rahmani et al. (2015) possesses lexically contrastive word prosody, whether due to stress or tone, so that any reliance on a "lexical memory" is not an option open to them.

The similar effects of stress and tone in the Dutch and Japanese accuracy scores in Rahmani et al. (2015) must not lead us to lose sight of the profoundly different character of tone from stress. Tones can form a great variety of subsystems in the phonologies of languages. There can be few or many of them and contrasts will vary in salience. And they could be lexical as well as morphological or syntactic ('grammatical'). Stress, by contrast, is usually taken to be the head of a constituent of the prosodic hierarchy, the foot, in which unstressed syllables may additionally occur in non-head positions (Selkirk, 1980; Hayes, 1995). Since all words are footed, and hence stressed, no stress contrasts are possible on monosyllables if a language has feet ("obligatoriness," Hyman, 2006). This is why the non-words in a stress-based SRT are disyllabic: stressed–unstressed or unstressed–stressed. At the same time, this makes it necessary to use monosyllabic contrasts in the case of tone, in order to guarantee tonal interpretations of the pitch contrasts. It is true that stress systems too vary across languages, for instance in the degree of exceptionality of stress locations. Moreover, stressed syllables may or may not have an intonational pitch accent, as in Germanic languages (*cf.* "primary stress," Domahs et al., 2008), and stress may correlate with syllable quantity or vowel reduction (Hayes, 1995). Such differences have not affected the results of SRTs much. In Peperkamp and Dupoux (2002), an experiment with six language groups, Polish, which has regular penultimate stress with few words having ultimate or antepenultimate stress, came out as intermediate between a stress-deaf and a non-stress-deaf group. Also, the categorical interaction between vowel quality and stress in European Portuguese explains why listeners are stress-deaf if they cannot rely on the vowel quality differences (Correia et al., 2015; Lu et al., 2018).

Because of the more varied complexity of lexical tone systems compared to stress systems, we may reasonably expect the results of a tonal SRT to be affected by relevant features of a language's phonology (Best, 2019). First, the number of monosyllabic tone melodies may vary from 2 to as many as 9 (e.g., Hyman, 2011). A high functional load of lexical pitch contrasts may well affect recall accuracy. Moreover, tone contrasts may be restricted to certain positions in the word, like the final syllable in Ma'ya (Remijsen, 2002) or a non-final syllable in Swedish (Riad, 2014: 182). This means that in addition to a simple discrete concept of lexical "tonality," that is, the presence of a pitch specification in the phonological form of at least some morphemes (Hyman, 2006), it will be necessary to test for effects of relative "tonality," that is, the complexity of lexical tone systems. Second, the choice of the pitch contrast in the experiment may favor participants that happen to have that contrast in their tonal grammar. We take this potential benefit to be independent of the lexical or intonational status of the pitch contrast. An experiment that intends to include this factor in its design, will need to test for a number of pitch contrasts, such that each of them fails to turn up in at least one language under investigation. Third, pitch contrasts vary in salience, that is, in the perceptual difference between the two contrasting pitch shapes. If sequences of less salient contrasts are harder to recall than contrasts with larger differences,

the size of the contrast will need to be included as a variable in our experiment.

We selected one unambiguously non-tonal language (Indonesian), one borderline case (Stockholm Swedish), and two unambiguously tonal languages (Taiwan Mandarin and Zhumadian Mandarin). The inclusion of two similar tone languages served as a sanity check, as it predicts that their scores will be quite similar as well as quite different from the non-tonal language. A heuristic element in our choice of languages is the ambiguous "semi-tonal" language, which might statistically side with either the non-tonal language or the tonal ones, or appear as a category in between.

Indonesian has neither tone nor stress on any syllable, whether word-based or phrase-based (Odé, 1994; Goedemans and van Zanten, 2007; Maskikit-Essed and Gussenhoven, 2016). The performance of the Indonesian participant group should provide a lower baseline. The language has an intonational contrast between a phrase-final rise, used in pre-final intonational phrases and in final interrogative phrases, and a rise–fall, used in final declarative phrases. The contrast between these right-edge melodies will show up in stated and questioned monosyllabic words. **Figure 1** shows this contrast as spoken by a 28-year-old male speaker from East Java. This pitch contrast is the main intonational contrast in the language and there may therefore be a fair bit of variation in the phonetic shapes.

Stockholm Swedish has a lexical tone contrast in non-final syllables with word stress, Accent 1 vs. Accent 2, as occurring in *anden* "the duck" and *anden* "the spirit," respectively. Accent 1 is a rise in the stressed syllable, followed by low pitch when occurring in the nuclear position, as illustrated by the solid line of an isolated pronunciation of the expression meaning "the duck" in **Figure 2**. Accent 2 has an early fall in the stressed syllable, which in the nuclear position is followed by a pitch peak in the phrase-final syllable, as shown by the dashed line for an isolated pronunciation of the expression meaning "the spirit" in **Figure 2**. Both have an intonational melody LHL%, which is preceded by a lexical H in the case of Accent 2, effectively shifting the intonational f0 peak onto the final syllable (Riad, 2014). Arguably, the different contours in the unstressed phrase-final syllables represent contrasting phonetic cues to the tone contrast on the penultimate syllable. However, such contextual cues abound in languages generally, so that we cannot interpret the phrase-final pitch difference as a contrast of the language, whether lexical or intonational.

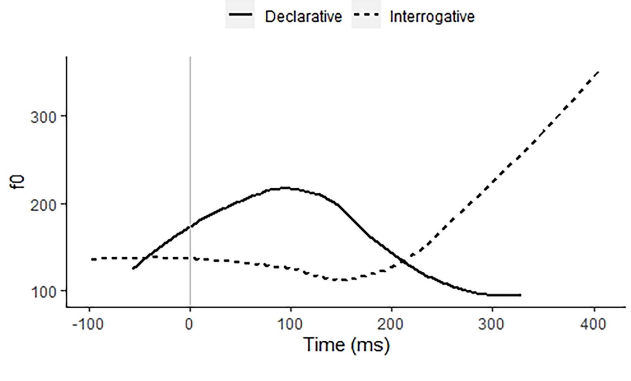

FIGURE 1 | f0 contours of declarative (solid line) and interrogative (dashed line) citation pronunciations of the monosyllabic word *gong* ("gong"), recorded by a 28-year-old male speaker of Standard Indonesian.

Zhumadian Mandarin, spoken in Henan Province, China, has four lexical tones, two rises, and two falls, which contrast for temporal alignment, leading to a late rise (Tone 1), a late fall (Tone 2), an early rise (Tone 3), and an early fall (Tone 4). The early rising Tone 3 tends to rise only a little, thus resembling Tone 1 of Standard Mandarin, while the late rising Tone 1 may sound like a final, dipping Tone 3 of Standard Mandarin (Gussenhoven and van de Ven, 2020). The language has a Fourth Tone Sandhi rule, changing 4+4 into 1+4, as well as toneless morphemes, that is, neutral tone. **Figure 3** presents examples of the four tones on the syllable /mae/. Younger speakers are bilingual with Standard Mandarin. Except in educational contexts, speakers use the Zhumadian dialect.

Taiwan Mandarin is a standard variety of Mandarin. It has four lexical tones, a high level tone, a rising tone, a low tone, and a high falling tone, Tones 1 to 4, respectively (**Figure 4**). In addition, it has the Third Tone Sandhi rule (3+3→2+3) as well as syllables with neutral tone, whose pitch contours are derivative from a preceding toned syllable. The most striking difference with Standard Chinese is the shorter duration of Tone 3, which typically lacks or significantly reduces the rising part in phrase-final position (Kubler, 1985; Fon and Chiang, 1999; Torgerson, 2005; Deng et al., 2006). Its tonal complexity is quite comparable to that of Zhumadian Mandarin.

MATERIALS AND METHODS

We included three-pitch contrasts in the experiment, EarlyFall vs. LateFall, EarlyRise vs. LateFall, and RiseFall vs. EarlyRise. None of these are pitch levels, which are likely to sound like a melody when occurring in a sequence, which would be more memorable than sequences of pitch shapes. In addition, we used a "phoneme" contrast of the type that has served as a control variable in SRT experiments (Peperkamp et al., 2010; Rahmani et al., 2015; Qin et al., 2017). A phonetically trained speaker of Dutch in his early 70s recorded each of these seven syllable types at least eight times in a sound-treated booth. Three tokens of each syllable type were selected that sounded natural and seemed good exemplars of the intended pitch shape. **Figure 5** displays these tokens for all five-pitch shapes figuring in these contrasts, all pronounced on the syllable [la], aligned at the onset-vowel boundary indicated by the gap in the figure, which corresponds to 0 ms in the signal. The phoneme contrast was between the syllables [ta] and [la], both pronounced with level midpitch. We avoided adjustments of the original durations, unlike Peperkamp et al. (2010), who drastically shortened the original recordings of disyllables. Largely depending on pitch shape, tones require a certain duration to produce (Xu and Sun, 2002) and shortened syllables may as a result sound distorted. Across pitch shape types, durations varied from

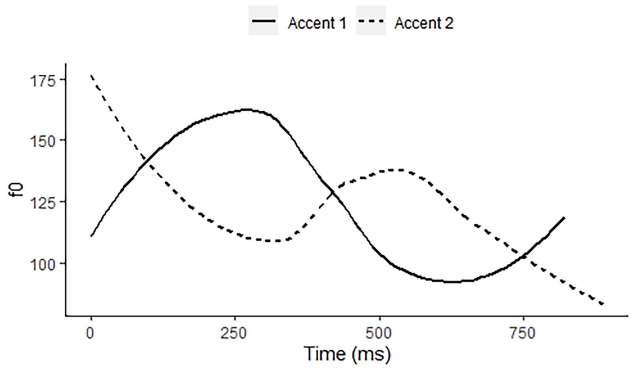

FIGURE 2 | f0 contours of citation pronunciations of Accent 1 on anden "the duck" (solid line) and Accent 2 on anden "the spirit" (dashed line) by a 60-year-old male speaker of Stockholm Swedish.

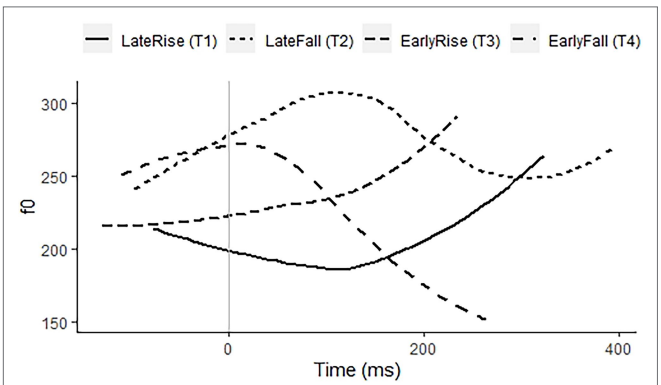

FIGURE 3 | f0 contours of citation pronunciations of a late rise/Tone 1 on 麥 "cereal," a late fall/Tone 2 on 埋 "bury," an early rise/Tone 3 on 買 "buy," and an early fall/Tone 4 on 賣 "sell," all with the segmental syllable /mae/, recorded by a 22-year-old female speaker of Zhumadian Mandarin.

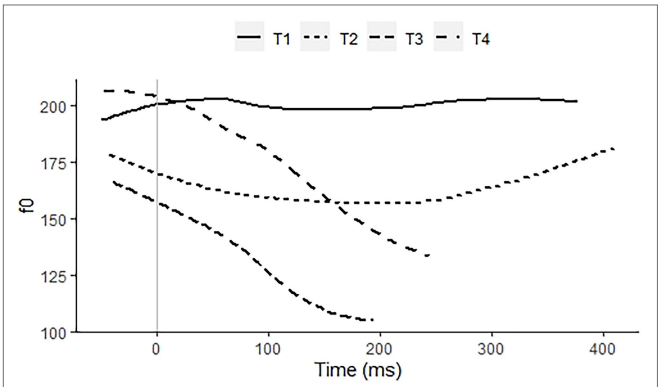

FIGURE 4 | f0 contours of citation pronunciations of a high level/Tone 1 on 媽 "mother," a rise/Tone 2 on 麻 "hemp," a low tone/Tone 3 on 馬 "horse," and a high falling/Tone 4 on 罵 "scold," all with the segmental syllable/ma/, recorded by a 40-year-old female speaker of Taiwan Mandarin.

430 ms for a token of the EarlyRise to 569 ms for a token of the RiseFall. The three tokens had very similar durations in three of the five-pitch shape types. Only the triplets for the

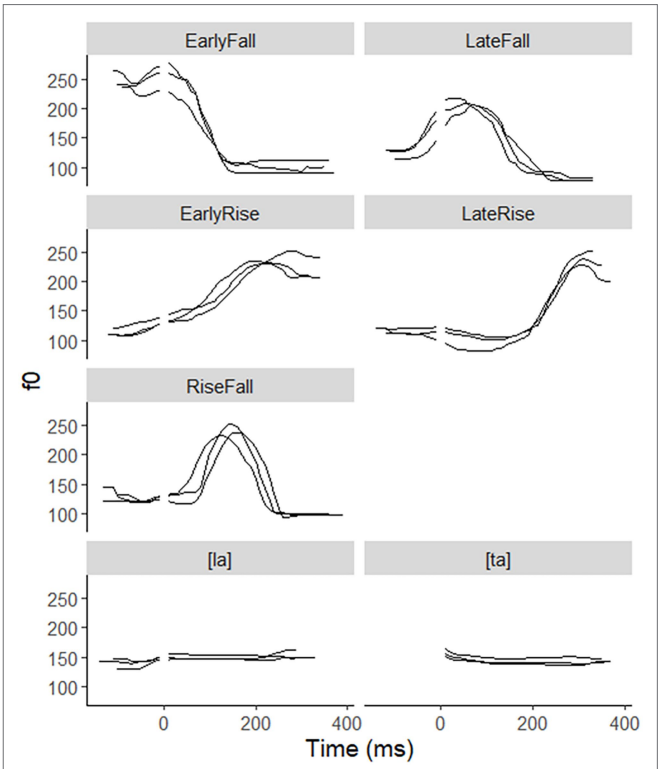

FIGURE 5 | f0 tracks of the three tokens for each of 7 syllables types, with the onset-vowel boundaries indicated by an interruption.

EarlyFall and the LateFall varied more noticeably, for which reason we standardized the three exemplars to the rounded mean duration in each triplet, 440 ms and 460 ms, respectively, using Praat (Boersma and Weenink, 1992–2020). **Figure 6** shows acoustic durations of all 15 pitch shape stimuli and the 6 stimuli for the phoneme contrast, for onset consonant and vowel separately; in the case of [ta], the burst duration is shown.

A number of independent variables were included in the analysis. SEX and APTITUDE were the two participant variables, of which APTITUDE was motivated by the expectation that participants may vary in their aptitude for carrying out an SRT. For this variable we used each participant's mean accuracy score on the phoneme contrast. Rather than controlling for pitch discrimination and categorization abilities, which have been shown to explain variation in pitch-related learning and identification tasks (*cf.* Sadakata and McQueen, 2014; Zhao and Kuhl, 2015; Bowles et al., 2016; Qin et al., 2021; Rhee et al., 2021), we intended to control for a more general ability to perform the experimental task of remembering sequences of tokens of two sound categories. Earlier research had taken this effect for granted, by subtracting phoneme accuracy scores from stress contrast scores (e.g., Peperkamp et al., 2010). We felt we needed to have a better understanding of the relation between the control and experimental contrasts in view of the prospect of continued research on languages with older populations of speakers.

Four language variables figured in our investigation, LEXICALITY, TONECOMPLEXITY, SALIENCE, and HAVECONTRAST.

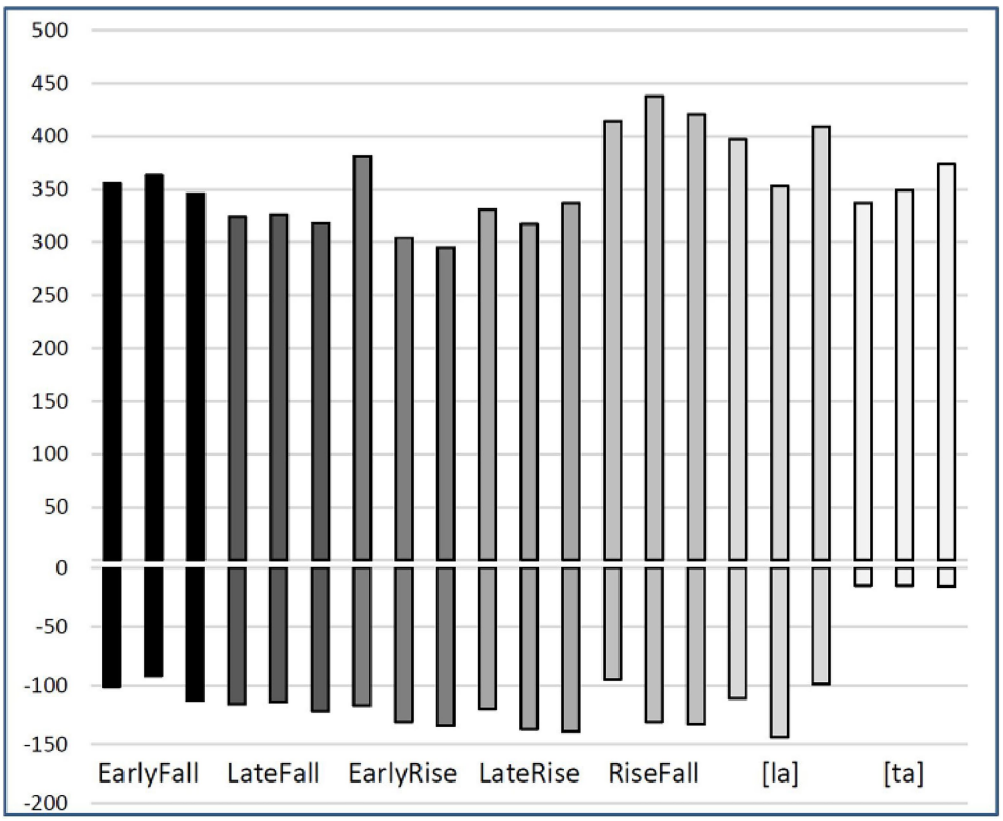

FIGURE 6 | Durations of onset [l] (negative bars) and rhyme [a] (positive bars) of the 27 stimuli in the experiment. For [ta], the negative bars give the positive VOTs. The value for the onset in [ta] is the burst and friction of the released [t].

Since our main hypothesis was that participants with tonal language backgrounds will outperform participants with non-tonal language backgrounds, we interpret LEXICALITY as a binary variable characterizing any language with a lexical marking of pitch as tonal (Hyman, 2006), which includes "semi-tonal" Swedish. While the distribution of the two Swedish tone categories is highly predictable from the phonology and morphology of words (Bruce, 1977: 18; Wetterlin et al., 2007; Riad, 2014: 183), there are exceptions, most obviously in disyllables with penultimat stress. For instance, many loan words have Accent 1, like *ketchup* and *solo*, in contrast to other words, like *senap* "mustard" and *pizza*, which have Accent 2. Moreover, in a priming experiment, Althaus et al. (2021) have shown that native speakers use the contrast in lexical access. Accordingly, only Indonesian was coded as −1 and the other three as 1 for this variable. At the same time, a gradient characterization of lexical tone complexity might provide a better predictor of accuracy scores than binary LEXICALITY, for which reason we coded the two Mandarin varieties as 4.0 for TONECOMPLEXITY, to reflect the number of tone categories. While Swedish has two tone categories, it has no tone contrast on monosyllabic words and hence not for the monosyllabic non-words in our experiment. We coded it as 0.5, while Indonesian was coded 0.0. Because LEXICALITY and TONECOMPLEXITY amount to discrete and gradual interpretations of a language's status as a tone language, we will not include both variables in the same analysis.

Our experiment involved pitch contrasts that obviously varied in salience. Because sequences of similar pitch shapes may be harder to recall than sequences of more different pitch shapes, we measured subjective phonetic differences among six-pitch shapes, one token of each of the five-pitch shapes in our experiment plus a FallRise, spoken by the same speaker, for the sake of symmetry in the set of pitch shapes to be measured. The 6 × 5 pairs were included in a Praat Multiple-Forced Choice experiment together with two filler pairs, presented in a per participant randomized order. Eight phonetically trained judges were asked to rate all pairs for phonetic distance on a 10-point scale, after listening to recordings of all six-pitch shapes and rating three trial pairs. Pearson's correlation coefficients between the scores of each judge and the mean score over all judges showed that the scores by two judges failed to reach significance at a 5% level. Of the other six, two judges had $r < 0.55$ and four $r > 0.83$.[1] They were native speakers of Dutch, English, Korean, and Mandarin

[1]In a methodologically comparable experiment with 40 participants, no effect of order of presentation within a pair was found (Fournier and Gussenhoven, 2010). In that experiment, the scores of all participants correlated with the mean scores over all judges, with a range of 0.56–0.88.

TABLE 1 | Mean subjective phonetic distances per pair of pitch shapes.

	EarlyFall	LateFall	EarlyRise	LateRise	RiseFall	FallRise
LateFall	**2.3**					
EarlyRise	9.5	**8.3**				
LateRise	9.5	8.7	4.3			
RiseFall	7.5	6.8	8.3	**8.7**		
FallRise	8.6	8.0	8.8	8.3	7.2	

Experimental contrasts are printed in bold.

TABLE 2 | Experimental pitch contrasts functioning as phonological contrasts.

Contrast	Indonesian	Swedish	Zhumadian Mandarin	Taiwan Mandarin
EarlyFall vs. LateFall	–1	–1	1	–1
LateFall vs. EarlyRise	–1	–1	1	1
LateRise vs. RiseFall	1	–1	–1	–1

(3), with ages ranging from 27 to 47. The native speaker of Korean grew up speaking tonal Gyeongsang Korean, but uses Standard Korean in virtually all domains. Their language backgrounds were otherwise evenly divided over tonal and non-tonal languages, which minimized language-specific biases (*cf.* Huang and Johnson, 2010). **Table 1** presents all scores, pooled over the two orders in each pair, which we used as the salience scores.

Finally, in order to be able to assess the extent to which the presence of a contrast in the participant's native language influences accuracy scores, we coded languages for HAVECONTRAST for each pitch contrast. When a language has a contrast in its lexical or postlexical phonology, it is coded 1 for that contrast, otherwise −1. For instance, Zhumadian Mandarin has a lexical contrast between an early aligning and a late-aligning fall, while the other three languages do not, entitling it to a 1 coding for that contrast (see **Table 2**). It also has a contrast between a late fall and an early rise, corresponding to our second experimental contrast. Taiwan Mandarin has a contrast between a fall and a late rise. Native speaker reactions suggest that the EarlyFall and the LateFall are equally good exemplars of the Taiwan Mandarin Fall, while the EarlyRise and the LateRise are both good exemplars of the Taiwan Mandarin Rise. We therefore also coded both Zhumadian and Taiwan Mandarin as 1 for the LateFall vs. EarlyRise contrast. Indonesian has an intonational contrast between a LateRise and a RiseFall, while the other three languages do not. Swedish lacks monosyllabic contrasts, so that it is harder to define the occurrence of our experimental contrast in the phonology of Swedish. Even if we were to interpret the f0 shapes of the first syllables as a RiseFall for Accent 1 and an EarlyFall for Accent 2 (see **Figure 2**), this would not correspond to any of the experimental pitch contrasts. Accordingly, all three contrasts are coded as −1 for Swedish.

We employed two sequence lengths for the two non-words, a 4-non-word and a 5-non-word sequence length, giving a binary variable SEQUENCELENGTH. Piloting with 6-non-word sequences made it clear that these were too difficult to deal with. In addition, we found that the task required a high level of concentration, which we felt put strict limits on the time participants could be asked to perform it. In a further attempt to make the task easier, we blocked the 4-non-word and presented these before moving on the block of 5-non-word sequences. Finally, GROUP and CONTRAST were the variables of central interest in the investigation. A summary of the independent variables introduced above appears in **Table 3**. Sequences of non-words avoided regular alternations (e.g., 1,212) and maximized the number of switch points (1 to 2, 2 to 1), following Rahmani et al. (2015), which led us to use 1211, 1221, 2112, 2122, 2212 and 1121 for 4-word sequences and 11221, 12112, 12212, 22112, 21221 and 21121 for 5-word sequences. With four contrasts and twice six sequences the total number of trials was 48. The total duration of the experiment was about 30 min.

We recruited minimally 20 participants for each language who were between 18 and 30 years old and attended or had attended institutes of tertiary education. **Table 4** lists the numbers per language split over the sexes, their age ranges, mean ages, and recruitment locations. We presented the experiment on a desktop computer with E-Prime 3.0 for the Zhumadian Mandarin participants and E-Prime 2.0 for the other participants (Schneider et al., 2012). Participants listened individually to the stimuli through headphones. Instructions were provided in English on the screen, supplemented with oral instructions in each native language. The experiment consisted of four blocks, one for each of the four contrasts with breaks in between, in a randomized order for each participant. Each block started with a training session. For the phoneme contrast, participants were trained to associate the syllable [la] with key "1" and [ta] with key "2," while for the three-pitch contrasts they were trained to associate [LateFall] with key "1" and [EarlyFall] with key "2," [LateFall] with key "1" and [EarlyRise] with key "2," and [LateRise] with key "1" and [RiseFall] with key "2."

Participants were told at the beginning of each block that they were going to learn two words in a foreign language. First, they heard all three tokens of one non-word with a "1" displayed on the screen, and then heard all three tokens of the other non-word with a "2" displayed on the screen. This cycle was repeated three times, exposing participants to 3 tokens x 2 non-words x 3 repetitions, or 18 non-words, before they proceeded to the second training stage, during which they heard each of the 6 tokens, together with a display of the corresponding key

TABLE 3 | Independent variables in the investigation.

Variable		Description
Experimental design	GROUP	Indonesian, Swedish, Zhumadian Mandarin, Taiwan Mandarin
	CONTRAST	EarlyFall vs. LateFall, LateFall vs. EarlyRise, LateRise vs. RiseFall, [la] vs. [ta]
	SEQUENCELENGTH	4-word sequence—1, 5-word sequence 1
Participant	SEX	Female—1, Male 1
	APTITUDE	Accuracy score [la]-[ta]
Linguistic structure	LEXICALITY	Indonesian—1, all other groups 1
	TONECOMPLEXITY	Indonesian 0.0, Swedish 0.5, Zhumadian Mandarin 4.0, Taiwan Mandarin 4.0
	HAVECONTRAST	See detailed coding in Table II.
	SALIENCE	EarlyFall vs. LateFall 2.3, LateFall vs. EarlyRise 8.3, LateRise vs. RiseFall 8.7

TABLE 4 | Participants in four language groups.

	N	Age range	Mean age	Location
Indonesian	10F, 10M	19–30	24.4	National Yang Ming Chiao Tung University (Hsinchu, Taiwan)
Swedish	11F, 10M	20–29	24.1	Stockholm University (Sweden)
Zhumadian M	15F, 10M	18–23	19.8	Huanghuai College (Zhumadian, China)
Taiwan M	10F, 10M	20–22	21.5	National Yang Ming Chiao Tung University (Hsinchu, Taiwan)

on the screen, in a random order. After they had indicated having learned the relevant two-way classification, participants moved on to an identification task in which they heard one of the six tokens in a contrast and were asked to respond by pressing "1" or "2." After each identification trial, they saw either "CORRECT!" or "INCORRECT!" on their screen for 800 ms as feedback. This procedure was repeated four times. The SRT proper was preceded by a warm-up block with six 3-word sequence trials. No feedback of any kind was given in the 4-sequence and 5-sequence experimental blocks. Ignoring the warm-up block, the experimental trials presented participants with all 48 stimulus pairs (6 sequences × 2 sequence lengths × 4 contrasts). Participants confirmed the completion of their response by pressing the ENTER key. The order of presentation of all sequences within all blocks was randomized per participant. Within each sequence, the non-words were randomly instantiated by one of the three tokens, while no token appeared more than once in a sequence.

Tokens were separated by 120-ms intervals in all sequences. Participants could only register their response after hearing a 1,600-ms recording of four piano chords, played 100 ms after the last token in a sequence. Its function was to reduce the ability of participants to rely on their acoustic memory, similar to that of the recording of "OK!" which has been used for SRTs with stress contrasts. Intervals between trials were 1,500 ms. No response was registered if its sequence length did not match that of the input sequence length.

RESULTS

Two analytical procedures were followed, after Peperkamp et al. (2010), one to answer the question what properties of the pitch contrast, the languages and the participants predict the accuracy scores and another to establish the differences between language groups and any interactions with the contrasts. Thus, we first report two multiple logistic regression analyses of the linguistic variables SALIENCE and HAVECONTRAST, together with the participant variables SEX and APTITUDE. In the first multiple logistic regression analysis, we included the binary variable LEXICALITY, while the gradient variable TONECOMPLEXITY was included in the second. We will next move on to building a mixed-effects model with the experimental design variables, including the phoneme control contrast [la] vs. [ta] (APTITUDE).

The results of the multiple logistic regression analysis on the accuracy scores for the three-pitch contrasts with SALIENCE, HAVECONTRAST, SEX, APTITUDE, and the binary variable LEXICALITY are given in **Table 5**. Significant HAVECONTRAST ($\beta = 0.29$, $p < 0.0001$) shows that participants generally have higher accuracy scores if some pitch difference they are judging is contrastive in their native language ("yes" $M = 0.63$ vs. "no" $M = 0.49$). SALIENCE ($\beta = 0.29$, $p < 0.0001$) indicates that the participants' performance relied to a large extent on how salient a specific contrast is. LEXICALITY ($\beta = 0.3$, $p < 0.0001$) also explained the accuracy results. Participants who speak a (semi-)tonal language ($M = 0.58$) outperformed Indonesian participants, whose native language lacks lexical tone ($M = 0.39$). Lastly, participants' performance on the three-pitch contrasts strongly depended on their scores for the phoneme contrast (APTITUDE, $\beta = 1.12$, $p < 0.0001$). The near-significant effect of SEX ($\beta = -0.07$, $p = 0.079$) weakly indicates that women ($M = 0.55$) performed better than men ($M = 0.52$). The model fit (r^2) is 0.24.

The results of the multiple logistic regression analysis with gradient TONECOMPLEXITY instead of LEXICALITY are given in **Table 6**. With a model fit (r^2) of 0.25, the explained variance is comparable, while the overall results for all identical variables are the same in the two analyses. The range of the accuracy means for TONECOMPLEXITY (0.39 to 0.63) is marginally wider than that for LEXICALITY (0.39 to 0.58) in the first analysis.

Next, two mixed-effects logistic regression analyses were performed on the accuracy scores to establish the effects of contrasts and language groups. The first focused on the tonally intermediate Swedish. With the Swedish participants and the phoneme contrast, [la] vs. [ta], as baselines, the regression model was fitted with CONTRAST * GROUP and SEQUENCELENGTH as variables, where CONTRAST has the three-pitch contrasts and the phoneme contrast as levels. In addition, the model included random intercepts for participant as well as by-participant random slopes for CONTRAST and SEQUENCELENGH. The second analysis was carried out to assess the degree of similarity between the two tonal languages, Taiwan vs. Zhumadian Mandarin. For this analysis, Taiwan Mandarin and the phoneme contrast were set as baselines, with the rest of the model structure remaining the same as that of the first. The analyses were run in R using the *lme4* package (Bates et al., 2015). The results of the two analyses are presented in **Tables 7** and **8**. **Figure 7** gives a box plot with accuracy means and per participant scatter plots.

The results of the first model show that the Swedish participants ($M = 0.88$) performed comparably at the phoneme contrast baseline with the Indonesian ($M = 0.86$) and Zhumadian Mandarin participants ($M = 0.87$), but marginally underperformed compared to the Taiwan Mandarin participants ($M = 0.93$; $\beta = 0.62$, $p = 0.06$). Swedish participants performed less well on the tonal contrasts than on the phoneme contrast (EarlyFall vs. LateFall ($M = 0.25$; $\beta = -3.51$, $p < 0.0001$), LateFall vs. EarlyRise ($M = 0.61$; $\beta = -1.79$, $p < 0.0001$) and the LateRise vs. RiseFall ($M = 0.60$; $\beta = -1.70$, $p < 0.0001$). Importantly, the Group–Contrast interactions indicate that the participants of the two tonal languages, Taiwan and Zhumadian Mandarin, outperformed Swedish participants on the LateFall vs. EarlyRise contrast (TM: $M = 0.90$, $\beta = 1.41$, $p < 0.001$; ZM: $M = 0.73$, $\beta = 0.76$, $p = 0.02$), while Swedish participants, in turn, outperformed non-tonal Indonesian participants on the same contrast ($M = 0.44$, $\beta = -0.92$, $p = 0.05$). Additionally, Zhumadian Mandarin participants performed better at the tonal contrast that is specific to their language, EarlyFall vs. LateFall, than the baseline Swedish participants ($M = 0.36$, $\beta = 0.74$, $p = 0.03$), while the results of the other two groups on this contrast were comparable to those of the Swedish group. Additionally, Taiwan Mandarin participants ($M = 0.86$) performed better on the LateRise vs. RiseFall contrast than the Swedish participants ($M = 0.60$; $\beta = 0.95$, $p = 0.02$). Finally, and unsurprisingly, 4-word sequences ($M = 0.69$) were responded to with higher accuracy than 5-word sequences ($M = 0.56$; $\beta = -0.42$, $p < 0.0001$).

The model with Taiwan Mandarin as the baseline shows that the Taiwan Mandarin group outperformed the Zhumadian Mandarin group on the phoneme baseline contrast ($\beta = -0.80$, $p = 0.01$); the difference with the Swedish group is just shy of significance. The low score for the Indonesian participants is not significantly different from the Taiwan Mandarin group, which is no doubt due to the wider spread of the scores by the Indonesian group compared to the concentration of the Taiwan Mandarin scores around 1 (**Figure 7**). Similar to the Swedish group, the Taiwan Mandarin group performed less well on the EarlyFall vs. LateFall ($M = 0.29$; $\beta = -3.93$, $p < 0.0001$) and the LateRise vs. RiseFall ($M = 0.86$; $\beta = -0.76$, $p = 0.07$) contrasts than on the phoneme contrast. Their performance on the LateFall vs. EarlyRise contrast, however, was as good as that on the phoneme contrast ($M = 0.90$; $\beta = -0.39$, $p = 0.32$). While the Taiwan Mandarin group still outperformed the non-tonal Indonesian and "semi-tonal" Swedish groups on the LateFall vs. EarlyRise and LateRise vs. RiseFall contrasts (Indonesian: $\beta = -2.32$, $p < 0.0001$; Swedish: $\beta = -1.41$, $p < 0.001$), the Zhumadian Mandarin group stood out on the Zhumadian-specific contrast, EarlyFall vs. LateFall ($M = 0.36$; $\beta = 1.16$, $p = 0.002$), the only contrast for which the Taiwan Mandarin group scored below Zhumadian Mandarin (see also **Figure 7**).

TABLE 5 | Results of a multiple logistic regression analysis with TONE COMPLEXITY as the tonality variable.

	$R^2 = 0.24$				
	B	**SE**	***z***	***p***	**Accuracy means**
Intercept	−2.709	0.235	−11.515	<0.0001	
HAVECONTRAST	0.288	0.042	6.809	<0.0001	no: 0.49; yes: 0.63
SALIENCE	0.290	0.014	20.533	<0.0001	2.3: 0.27; 8.3: 0.67; 8.7: 0.67
APTITUDE	1.123	0.286	3.927	<0.0001	
LEXICALITY	0.302	0.066	4.61	<0.0001	−1: 0.39; 1: 0.58
SEX	−0.071	0.04	−1.751	0.079	female: 0.55; male: 0.52

TABLE 6 | Results of a multiple logistic regression analysis with TONE COMPLEXITY as the tonality variable.

	$R^2 = 0.25$				
	B	**SE**	***z***	***p***	**Accuracy means**
Intercept	−3.17	0.207	−15.289	<0.0001	
HAVECONTRAST	0.181	0.046	3.954	<0.0001	
SALIENCE	0.296	0.014	20.668	<0.0001	
APTITUDE	1.374	0.233	5.905	<0.0001	
TONECOMPLEXITY	0.165	0.025	6.533	<0.0001	0.0: 0.39; 0.5: 0.49; 4.0: 0.63
SEX	−0.07	0.04	−1.74	0.081	

TABLE 7 | Results of mixed-effects logistic regression analysis with Swedish and [la] vs. [ta] as baselines.

	$R^2 = 0.47$			
	B	**SE**	***z***	***p***
Intercept	2.318	0.300	7.716	<0.0001
GroupIndonesian	0.025	0.428	0.059	0.953
GroupZhumadian M.	−0.187	0.274	−0.681	0.496
GroupTaiwan M.	**0.615**	**0.327**	**1.882**	**0.060**
ContrastEarlyFall* vs. *LateFall	**−3.510**	**0.321**	**−10.925**	**<0.0001**
ContrastLateFall* vs. *EarlyRise	**−1.794**	**0.309**	**−5.802**	**<0.0001**
ContrastLateRise* vs. *RiseFall	**−1.703**	**0.354**	**−4.819**	**<0.0001**
Sequence	**−0.421**	**0.044**	**−9.664**	**<0.0001**
GroupIndonesian:ContrastEarlyFall vs. *LateFall*	−0.718	0.473	−1.520	0.129
GroupZhumadian M.:ContrastEarlyFall* vs. *LateFall	**0.738**	**0.337**	**2.193**	**0.028**
GroupTaiwan M.:ContrastEarlyFall vs. *LateFall*	−0.417	0.389	−1.072	0.284
GroupIndonesian:ContrastLateFall* vs. *EarlyRise	**−0.917**	**0.458**	**−2.003**	**0.045**
GroupZhumadian M.:ContrastLateFall* vs. *EarlyRise	**0.761**	**0.337**	**2.258**	**0.024**
GroupTaiwan M.:ContrastLateFall* vs. *EarlyRise	**1.407**	**0.420**	**3.345**	**0.001**
GroupIndonesian:ContrastLateRise vs. *RiseFall*	−0.420	0.510	−0.823	0.410
GroupZhumadian M.:ContrastLateRise vs. *RiseFall*	0.403	0.334	1.206	0.228
GroupTaiwan M.:ContrastLateRise* vs. *RiseFall	**0.948**	**0.405**	**2.342**	**0.019**

Significant results are presented in bold.

DISCUSSION

There are three main results of our experiment on the sequence recall of pitch shapes with Indonesian, Swedish, and Mandarin participants.

1. Accuracy scores were positively influenced by (i) similarities between experimental pitch contrasts and phonological contrasts in the languages, (ii) the phonetic salience of the experimental pitch contrast, and (iii) the participant's aptitude for the experimental task as measured by the score on the phoneme contrast.
2. On one contrast, LateFall vs. EarlyRise, the Swedish group distinguished themselves as intermediate by outperforming the Indonesian group and being outperformed by the two Mandarin groups, with the two Mandarin groups not differing among themselves.
3. On none of the three-pitch contrasts did semi-tonal Swedish participants and the two tonal Mandarin groups outperform the non-tonal Indonesian group without differing among themselves.

We discuss these three findings in this order below.

Dependence of Tone Contrast Sequence Recall Accuracy Scores on Other Factors

Without a doubt, the linguistic effects of our first finding will show up in similar experiments performed with different selections of languages. Given the small size of our experiment, we cannot be confident that the effect sizes will be preserved proportionally in experiments with different sets of pitch contrasts and languages, but our results do show that a tonal SRT will need to address the effects of linguistic properties to a larger extent than a stress-based SRT (*cf.* Best, 2019). Despite the cross-linguistic variation in the distribution of stressed syllables within words outlined in Peperkamp and Dupoux (2002), the cross-linguistic variation in tone systems is larger than that of stress.

The effect of the general ability of participants to perform an SRT, as measured by the accuracy scores of the phoneme contrast (APTITUDE), turned up among four groups of participants with similar age ranges and levels of education. This suggests that for older participants, this task may be more challenging and hence likely to produce lower accuracy scores compared to our participants. Less demanding versions of this experimental task may therefore need to be explored with older participants. As far as we are aware, this is the first time that an SRT aptitude effect has shown up. Rahmani et al. (2015) ignored the phoneme contrast for not being significantly different between language groups. In Peperkamp et al. (2010), the dependent variable was the difference between the accuracy scores for the phoneme contrast and the stress contrast, on the assumption that this effect will exist in absolute terms, while excluding participants showing poor performance from the analysis, resulting in a significant data loss. By including the phoneme contrast scores as a variable in our multiple regression analyses and the model analyses, we were able to retain all participants in the experiment so as to closely model their performance. Various components of aptitude have been addressed in more recent studies as a variable that could potentially modulate tone perception, as in Bowles et al. (2016) and Qin et al., (2021).

The effect of the existence of an experimental pitch contrast in a language's phonology (HAVECONTRAST) is apparent from

TABLE 8 | Results of mixed-effects logistic regression analysis with Taiwanese Mandarin and [la] vs. [ta] as baselines.

	$R^2 = 0.47$			
	B	**SE**	***z***	***p***
Intercept	2.934	0.341	8.604	<0.0001
GroupZhumadian M.	**−0.802**	**0.316**	**−2.542**	**0.011**
GroupSwedish	**−0.615**	**0.327**	**−1.883**	**0.060**
GroepIndonesian	−0.590	0.456	−1.293	0.196
ContrastEarlyFall vs. LateFall	**−3.926**	**0.358**	**−10.957**	**<0.0001**
ContrastLateFall vs. *EarlyRise*	−0.387	0.393	−0.987	0.324
ContrastLateRise vs. *RiseFall*	−0.756	0.416	−1.817	0.069
Sequence	**−0.421**	**0.044**	**−9.664**	**<0.0001**
GroupZhumadian M.:ContrastEarlyFall vs. LateFall	**1.155**	**0.370**	**3.119**	**0.002**
GroupSwedish:ContrastEarlyFall vs. *LateFall*	0.417	0.389	1.072	0.284
GroupIndonesian:ContrastEarlyFall vs. *LateFall*	−0.302	0.498	−0.606	0.544
GroupZhumadian M.:ContrastLateFall vs. *EarlyRise*	−0.646	0.411	−1.570	0.116
GroupSwedish:ContrastLateFall vs. EarlyRise	**−1.407**	**0.420**	**−3.347**	**0.001**
GroupIndonesian:ContrastLateFall vs. EarlyRise	**−2.324**	**0.518**	**−4.485**	**< 0.0001**
GroupZhumadian M.:ContrastLateRise vs. *RiseFall*	−0.545	0.395	−1.380	0.168
GroupSwedish:ContrastLateRise vs. RiseFall	**−0.948**	**0.404**	**−2.343**	**0.019**
GroupIndonesian:ContrastLateRise vs. RiseFall	**−1.368**	**0.557**	**−2.455**	**0.014**

Significant results are presented in bold.

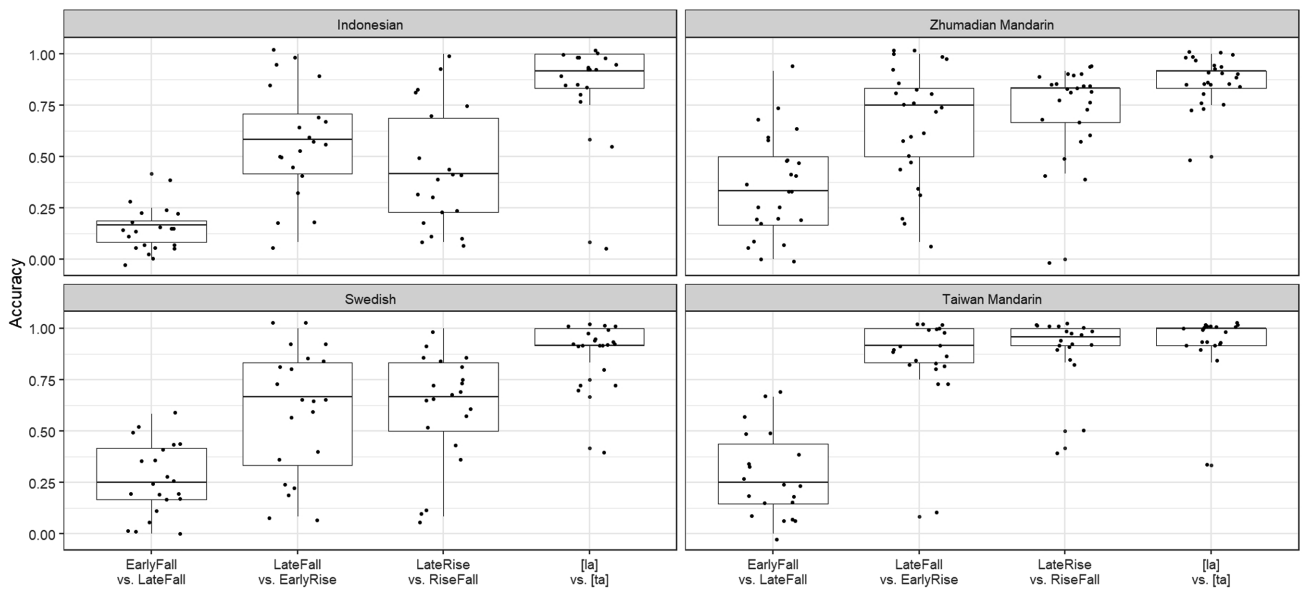

FIGURE 7 | Boxplots and scatterplots for four contrasts and four language groups.

the interactions between the pitch contrasts and the language groups in the mixed-effects models. The Zhumadian group, whose language is the only one to have a temporal alignment contrast for falls, outperformed both the Swedish and Taiwan Mandarin groups on the EarlyFall vs. LateFall contrast, in addition to the low-scoring Indonesian group. The three non-Zhumadian groups did not differ significantly from each other, as shown by the lack of any interaction between Indonesian and the EarlyFall vs. LateFall contrast in either analysis (**Tables 7** and **8**). The effect of contrast salience (SALIENCE) was most

clearly in evidence in the overall lower scores for the EarlyFall vs. LateFall contrast compared to the other two pitch contrasts.

Three Typological Groups?

Our second finding was that both Mandarin groups outperformed the Indonesian and Swedish groups on the LateFall vs. EarlyRise contrast, with the Indonesian group scoring below the Swedish group. If we interpret the contrast between rising and falling pitch to be prototypical, the pattern Indonesian < Swedish < Zhumadian and Taiwan Mandarin suggests a three-way distinction between atonal, semi-tonal, and tonal languages. If this result were to be replicated with other mixes of languages, it would imply that a binary diagnostic is unlikely to emerge from a tone-based SRT with a broad typological mix of languages. In turn, this might put experiments with small numbers of languages that have yielded significant results between tonal and non-tonal languages in a different perspective, in the sense that they may represent values on a tone/non-tone continuum rather than as values of a binary variable.

Testing Varieties of the Same Language

Turning the above conclusion around so as to adopt a positive perspective, we might expect tonal and non-tonal varieties of the same language that otherwise have few differences between them to be consistently distinguishable with the help of a tonal SRT. Such languages include Japanese, Korean, Swedish/Norwegian, Franconian varieties of Dutch and German, and Serbian/Croatian (van der Hulst et al., 2011; Gussenhoven and Chen, 2020). Importantly, it is in such cases that the tonal nature of languages has been debated, most notably with respect to two properties, one distributional and the other representational. The first is exemplified by Tokyo Japanese and Northern Bizkayan Basque, which have been characterized as "pitch accent languages," a distinct type by the side of tonal and non-tonal languages. Dominant characterizations of this group indicate the restriction of contrastive tone in a single location of the word or word-like domain. Hyman (2006, 2009) has signaled the absence of a clear definition, in particular that of the demarcation line with tone languages proper. Thus, the single location could be "fixed," like the penultimate syllable of Lekeitio Basque, be restricted to the non-final stressed syllable, as in Swedish, or to one of two syllables at a word edge, as in Kagoshima Japanese and Barasana, or be lexically specified, as in Tokyo Japanese (Elordieta, 1998; Gomez-Imbert and Kenstowicz, 2000; Hualde, 2012; Jun and Kubozono, 2020). Also, there may be two locations for a tone contrast, one at the beginning and one toward the end, as in Osaka and Ibukujima Japanese (Pierrehumbert and Beckman, 1988; Uwano, 1999), while the contrastive tone could be privative, as in the above varieties of Japanese, or represent a contrast between two tone melodies, as in Barasana (*cf.* Hualde, 2012). The other controversy concerns the issue whether surface tone contrasts in varieties of Swedish/Norwegian and Franconian are due to underlying tones (e.g., Bruce, 1977; Riad, 2014; Gussenhoven and Peters, 2019) or to differences in underlying foot structure which generate the different surface tone structures (e.g., Köhnlein, 2011, 2016, 2017; Hermans, 2012; Morén-Duolljá, 2013; Kehrein, 2018). Future explorations of our tone-based SRT might therefore fruitfully compare non-tonal and putatively tonal varieties of the same language.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/uxv4j/?view_only=86c8981b6c9c46c38fe2c2900afd4bcc.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee for Human Subject Protection of National Yang Ming Chiao Tung University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CG: conceptualization, methodology, data curation, supervision, and writing—original draft. Y-AL: data curation, formal analysis, funding acquisition, project administration, resources, visualizations, software, Taiwan Mandarin and Indonesian experiments, and writing—review and editing. S-IL-K: methodology, formal analysis, and writing—review and editing. CL: resources, Zhumadian Mandarin experiment, and writing—review and editing. HR: resources. TR: writing—review and editing. HZ: Swedish experiment and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

The work by CG was supported by MOST-208-2811-H-009-500 (Ministry of Science and Technology, Taiwan) awarded to Ho-hsien Pan. Taiwan Mandarin and Indonesian participants were run using MOST 109-2410-H-009-048 granted to Y-AL.

REFERENCES

Althaus, N., Wetterlin, A., and Lahiri, A. (2021). Features of low functional load in mono- and bilinguals' lexical access: evidence from Swedish tonal accent. *Phonetica* 78, 175–199. doi: 10.1515/phon-2021-2002

Baddeley, A. (2010). Working memory. *Current Biology* 20, R136–140. doi: 10.1016/j.cub.2009.12.014

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Best, C. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: considerations for tone perception research. *Front. Psychol.* 10:364. doi: 10.3389/fpsyg.2019.00364

Boersma, P., and Weenink, D. (1992-2020). Doing phonetics by computer. Available at: www.praat.org

Bowles, A. R., Chang, C. B., and Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Lang. Learn.* 66, 774–808. doi: 10.1111/lang.12159

Bruce, G. (1977). *Swedish Word Accent in Sentence Perspective*. Lund: Gleerup.

Correia, S., Butler, J., Vigário, M., and Frota, S. (2015). A stress "deafness" effect in European Portuguese. *Lang. Speech* 58, 48–67. doi: 10.1177/0023830914565193

Deng, D., Shi, F., and Lu, S. (2006). The contrast on tone between Putonghua and Taiwan Mandarin. *Acta Acustica* 31, 536–541.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25, 1–36. doi: 10.1017/S0952675708001383

Dupoux, E., Pallier, C., Sebastián, N., and Mehler, J. (1997). A destressing 'deafness' in French? *J. Memory Lang.* 36, 406–421. doi: 10.1006/jmla.1996.2500

Dupoux, E., Peperkamp, S., and Sabastián-Gallés, N. (2001). A robust method to study stress 'deafness'. *J. Acoust. Soc. Am.* 110, 1606–1618. doi: 10.1121/1.1380437

Dupoux, E., Sebastián-Gallés, N., Navarete, E., and Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition* 106, 682–706. doi: 10.1016/j.cognition.2007.04.001

Elordieta, E. (1998). Intonation in a pitch accent variety of Basque. *ASJU: Int. J. Basque Ling. Philology* 32, 511–569.

Fon, J., and Chiang, W.-Y. (1999). What does Chao have to say about tones? A case study of Taiwan Mandarin/赵氏声调系统与声学之联结及量化–以台湾地区国语为例. *J. Chin. Ling.* 27, 13–37.

Fournier, R., and Gussenhoven, C. (2010). Measuring phonetic salience and perceptual distinctiveness: the lexical tone contrast of Venlo Dutch. *Revista Diadorim: Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro*, 12. Available at: http://www.revistadiadorim.letras.ufrj.br (Accessed September 24, 2021).

Goedemans, R., and van Zanten, E. (2007). "Stress and accent in Indonesian," in *Prosody in Indonesian Languages*. eds. V. J. van Heuven and E. van Zanten (Utrecht: LOT), 35–62.

Gomez-Imbert, E., and Kenstowicz, M. (2000). Barasana tone and accent. *Int. J. Am. Ling.* 66, 419–463. doi: 10.1086/466437

Gooden, S. (2022). Intonation and prosody in creole languages: an evolving typology. *Annu. Rev. Ling.* 8, 343–364. doi: 10.1146/annurev-linguistics-031120-124320

Gussenhoven, C., and Chen, A. (2020). *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press.

Gussenhoven, C., and Peters, J. (2019). Franconian tones fare better as tones than as feet: a reply to Köhnlein (2016). *Phonology* 36, 497–530. doi: 10.1017/S095267571900023X

Gussenhoven, C., and van de Ven, M. (2020). Categorical perception of lexical tone contrasts and gradient perception of the statement-question intonation contrast in Zhumadian Mandarin. *Lang. Cogn.* 12, 614–648. doi: 10.1017/langcog.2020.14

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: Chicago University Press.

Hermans, B. (2012). "The phonological representation of the Limburgian tonal accents," in *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. eds. B. Botma and R. Noske (Berlin: Mouton de Gruyter), 223–239.

Hualde, J. I. (2012). Two Basque accentual systems and word-prosodic typology. *Lingua* 122, 1335–1351. doi: 10.1016/j.lingua.2012.05.003

Huang, T., and Johnson, K. (2010). Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica* 10, 243–267. doi: 10.1159/000327392

Hyman, L. M. (2006). Word prosodic typology. *Phonology* 23, 225–257. doi: 10.1017/S0952675706000893

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch accent. *Lang. Sci.* 31, 213–238. doi: 10.1016/j.langsci.2008.12.007

Hyman, L. M. (2011). "Tone: is it different?" in *The Handbook of Phonological Theory. 2nd Edn*. eds. J. A. Goldsmith, J. Riggle and A. Yu (Malden: Wiley Blackwell), 197–239.

Hyman, L. M. (2016). "Lexical vs. Grammatical Tone: Sorting out the Differences." in *Proceedings of the 5th International Symposium on Tonal Aspects of Languages* (Buffalo, NY: TAL 2016), 6-11.

Jun, S.-A., and Kubozono, H. (2020). "Asian Pacific Rim," in *The Oxford Handbook of Language Prosody*. eds. C. Gussenhoven and A. Chen, (Oxford: Oxford University Press), 355–369.

Kehrein, W. (2018). "There's no tone in Cologne: against tone-segment interactions in Franconian," in *Segmental Structure and Tone*. eds. W. Kehrein, B. Köhnlein, P. Boersma and M. van Oostendorp (Berlin: Mouton de Gruyter), 147–194.

Kehrein, W., Köhnlein, B., Boersma, P., and van Oostendorp, M. (2017). *Segmental Structure and Tone*. Berlin: Mouton de Gruyter, 147–194.

Köhnlein, B. (2011). Rule reversal revisited: synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach. PhD dissertation: Leiden. LOT Dissertation series 274.

Köhnlein, B. (2016). Contrastive foot structure in Franconian tone-accent dialects. *Phonology* 33, 87–123. doi: 10.1017/S095267571600004X

Köhnlein, B. (2017). "Synchronic alternations between monophthongs and diphthongs in Franconian tone accent dialects: a metrical approach," in *Segmental Structure and tone*. eds. W. Kehrein, B. Köhnlein, P. Boersma and M. van Oostendorp (Berlin: Mouton de Gruyter), 211–235.

Kubler, C. C. (1985). The influence of southern min on the mandarin of Taiwan. *Anthropol. Ling.* 27, 156–176.

Lau, J. C. Y., Xie, Z., Chandrasekaran, B., and Wong, P. C. M. (2020). "Cortical and subcortical processing of linguistic pitch patterns," in *The Oxford Handbook of Language Prosody*. eds. C. Gussenhoven and A. Chen (Oxford: Oxford University Press), 499–508.

Lu, S., Vigário, M., Correia, S., Jerónimo, R., and Frota, S. (2018). Revisiting stress "deafness" in European Portuguese: a behavioral and ERP study. *Front. Psychol.* 9:2486. doi: 10.3389/fpsyg.2018.02486

Maskikit-Essed, R., and Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology* 33, 353–389. doi: 10.1017/S0952675716000154

Morén-Duolljá, B. (2013). The prosody of Swedish underived nouns: no lexical tones required. *Nordlyd* 40, 196–248. doi: 10.7557/12.2506

Odé, C. (1994). "On the perception of prominence in Indonesian," in *Experimental Studies of Indonesian Prosody*. eds. C. Odé, V. J. van Heuven and E. van Zanten (Leiden University: Rijksuniversiteit te Leiden, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië), 27–107.

Peperkamp, S. (2004). Lexical exceptions in stress systems: arguments from early language acquisition and adult speech perception. *Language* 80, 98–126. doi: 10.1353/lan.2004.0035

Peperkamp, S., and Dupoux, E. (2002). "A typological study of stress 'deafness'" in *Laboratory Phonology 7*. eds. C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 203–240.

Peperkamp, S., Vendalin, I., and Dupoux, E. (2010). Perception of predictable stress: a cross-linguistic investigation. *J. Phon.* 38, 422–430. doi: 10.1016/j.wocn.2010.04.001

Pierrehumbert, J. B., and Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Qin, Z., Chien, Y.-F., and Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Appl. Psycholinguist.* 38, 541–570. doi: 10.1017/S0142716416000321

Qin, Z., Zhang, C., and Wang, W. S.-Y. (2021). The effect of mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *J. Acoust. Soc. Am.* 149, 435–446. doi: 10.1121/10.0003330

Rahmani, H., Rietveld, T., and Gussenhoven, C. (2015). Stress "deafness" reveals absence of lexical marking of stress or tone in the adult grammar. *PLoS One* 10:e0143968. doi: 10.1371/journal.pone.0143968

Remijsen, B. (2002). "Lexically contrastive stress accent and lexical tone in Ma'ya," in *Laboratory Phonology. Vol. 7*. eds. C. Gussenhoven and N. Warner (Berlin/New York: Mouton de Gruyter), 585–614.

Rhee, N., Chen, A., and Kuang, J. (2021). Musicality and age interaction in tone development. *Front. Neurosci.* 16:804042. doi: 10.3389/fnins.2022.804042

Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.

Sadakata, M., and McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Front. Psychol.* 5:1318. doi: 10.3389/fpsyg.2014.01318

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Ling. Inq.* 11, 563–605.

Steien, G. B., and Yakpo, K. (2020). Romancing with tone: on the outcomes of prosodic contact. *Language* 96, 1–41. doi: 10.1353/lan.2020.0000

Torgerson, R. C. (2005). A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles. A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of Three native speech styles. PhD dissertation. Provo, UT: Brigham Young University.

Uwano, Z. (1999). "Classification of Japanese accent systems," in *Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*. ed. S. Kaji (Tokyo: ILCAA, Tokyo University of Foreign Studies), 151–186.

van der Hulst, H., Goedemans, R., and van Zanten, E. (2011). *A Survey of Word Accentual Patterns in the Languages of the World*. Berlin: Mouton de Gruyter.

van Heuven, V. J., and Turk, A. (2020). "Phonetic correlates of word and sentence stress," in *The Oxford Handbook of Language Prosody*. eds. C. Gussenhoven and A. Chen (Oxford: Oxford University Press), 1501–1565.

Wetterlin, A., Jönsson-Steiner, E., and Lahiri, A. (2007). "Tones and loans in the history of Scandinavian," in *Tones and Tunes. Volume 1: Typological Studies in Word and Sentence Prosody*. eds. T. Riad and C. Gussenhoven, (Berlin: Mouton de Gruyter), 353–375.

Xu, Y., and Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *J. Acoust. Soc. Am.* 111, 1399–1413. doi: 10.1121/1.1445789

Zhao, T. C., and Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *J. Acoust. Soc. Am.* 137, 1452–1463. doi: 10.1121/1.4913457

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.