

Categorical perception of lexical tone contrasts and gradient perception of the statement–question intonation contrast in Zhumadian Mandarin*

CARLOS GUSSENHOVEN 

*Department of Foreign Languages and Literatures, National Chiao Tung University, Hsinchu, Taiwan, and
Centre for Language Studies, Radboud University, Nijmegen, The Netherlands*

AND

MARCO VAN DE VEN 

Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

*(Received 23 October 2019 – Revised 24 March 2020 – Accepted 26 March 2020 –
First published online 28 April 2020)*

ABSTRACT

We intended to establish if two lexical tone contrasts in Zhumadian Mandarin, one between early and late aligned falls and another between early and late aligned rises, are perceived categorically, while the difference between declarative and interrogative pronunciations of these four tones is perceived gradiently. Presenting stimuli from 7-point acoustic

[*] We thank Naodiu Li for serving as an informant of her language during a class the first author taught at Nanjing University in the fall of 2012, and for recording our materials. We are inestimably indebted to Minghui Wu, who arranged the contacts with Huanghuai College and No 1 Senior Middle School in Quesan, and Yuanyuan Zhang, who helped run the experiments in Zhumadian in the summer of 2015, as well as to Irham Aladist, Ad Foolen, and Irham Rofiq, with whom we ran the Indonesian experiments in the spring of 2017 at the Maulana Malik Ibrahim State Islamic University in Malang. Polina Drozdova, Paula Escudero, Janet van Hell, Joop Kerkhoff, Anne Voncken, Lei Wang, and Patrick van de Zande provided technical assistance, while John Kingston gave welcome advice during the design stage of the experiments. We acknowledge the comments made by members of the audiences after presentations of this research at LabPhon16 (University of Lisbon, 19–22 June 2018), ICPP5 (NINJAL, Tokyo, 26–28 October 2018), TIE2018 (Gothenburg, 11–13 October 2018), Academia Sinica (Taipei, 11 September 2018), National Chiao Tung University (Hsinchu, 13 September 2018), Stanford University (30 November 2018), UC Berkeley (3 December 2018), and the TIN-dag meeting (Utrecht, 2 February 2019). We are particularly grateful for the discussion with Wendy Sandler, for statistical advice from Louis ten Bosch, and for the comments by our reviewers and editor Bodo Winter on an earlier version of this paper.

continua between tones and between intonations, we used an identification task and a discrimination task with an experimental group of native listeners and a control group of Indonesian listeners, whose language employs none of the differences within either the falling or the rising pitch contours in its phonology. Only the lexical condition as perceived by the experimental group yielded sigmoid identification functions and a heightened discriminatory sensitivity around the midpoint of continua. The intonational condition in the native group and both conditions in the control group yielded gradient identification functions and smaller, reverse effects of the continuum midpoints in the discrimination task. The results are interpreted to mean that sentence modality contrasts can be expressed gradiently, but that lexical tone differences are represented phonologically, and hence are perceived categorically, despite low phonetic salience of the contrast. This conclusion challenges assumptions about the relation between linguistic functions and linguistic structures.



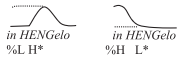

KEYWORDS: categorical perception, Zhumadian Mandarin, question intonation, tone alignment contrast, Indonesian

1. Introduction

A long-standing as well as controversial issue in language communication concerns the dichotomy between grammatically encoded morphemes and a range of systematic phonetic patterns that reflect further aspects, like word-specific phonetics, speaker attitudes, and indications of sound change, as well as the speaker's age, gender, and social class. This issue has been raised many times in studies of intonation in particular, and has been formulated as one between discreteness and gradience, or typically equivalently, between 'language' in a narrower structural sense and 'paralanguage'. In the words of Ladd (2008 [1996], p. 37), the difference "resides in the *quantal* or *categorical* structure of linguistic signalling, and the *scalar* or *gradient* nature of paralanguage". In this view, intonation in a broad sense is partly represented as tonally encoded discourse-marking morphemes and partly as meaningful pitch features that adjust the pitch contour arising from the phonological structure, with the degree of the phonetic adjustments being commensurate with the forcefulness with which meanings are expressed.

While distinguishing between these two aspects of the speaker's knowledge of their language's sound structure is widely seen as a central analytical task (Ladd, 2008 [1996], p. 39; Gussenhoven, 2004, Ch. 5; Prieto, 2012), this must not lead us to think that discrete features are somehow more central to linguistics than gradient ones, a view that is likely to be bound up with the belief that linguistic meanings are necessarily morphemic, i.e., represented phonologically, and that gradience is reserved for expressing the speaker's

TABLE 1. *Functions of gradient and discrete uses of pitch (references in text) with simplified illustrative pitch contours on brief examples. Capitalized syllables are accented, H* stands for a focus-marking pitch accent, %T for a phrase-initial boundary tone, which in the positive emotion (solid) is %L if the next tone is H* and %H if the next tone is L*. The solid pitch contours in the ‘Gradient’ column indicate a raised register.*

	<i>Discrete</i>	<i>Gradient</i>
<i>Informational</i>	H* to express ‘newness’ (European languages) 	Raised register to express interrogativity (Swedish) 
<i>Affective</i>	Polar %T to express positive emotion (Dutch) 	Raised register to express friendliness (English) 

emotions or evaluations in relation to the linguistic message (e.g., Weitz, 1979, p. 94). Intonation research has shown that this position is false. Bruce (1977, p. 137) found that sentence modalities in Stockholm Swedish involve adjustments of the f0 baseline, implying that the register raising as used for questions is gradient. That is, gradience in intonation is not confined to ‘affective’ meanings (meanings applying to the speaker), and may include ‘informational’ meanings (meanings applying to the message) (Grice & Baumann, 2007; Gussenhoven, 2004; Prieto, 2015). In Table 1, ‘informational’ is used for what has also been called ‘logical’, from Monroy (2005), credited to Bally (1935), and, more widely, ‘propositional’ or ‘linguistic’. Unsurprisingly, informational meanings are vastly more likely to be morphemic than affective meanings, as exemplified by widely reported focus-marking functions of pitch accents (e.g., Ladd 2008 [1996], p. 213). Similarly, gradiently expressed affective meanings, like register raising to express a positive disposition to the hearer (Chen, Gussenhoven, & Rietveld, 2004), is commonplace. But the meaning and form dimensions are ultimately independent: gradient emphasis may occur in languages that lack formal means of encoding focus, just as interrogativity may be expressed gradiently in intonation. As for discretely expressed affective

meaning, a claim for a morpheme with affective meaning has been made for a polar boundary tone in Dutch, which expresses positive affective meanings if its value is opposite to the upcoming tone of a pitch accent (Grabe, Gussenhoven, Haan, Marsi, & Post, 1998; Gussenhoven, 2004, p. 88) (see Table 1).

Both gradient and discrete inputs to the speech signal are captured in the way Roessig, Mücke, and Grice (2019) model variation across as well as within intonational categories in German. In their dynamical system, stable states ('attractors') reflect phonological categories, such that within-category and between-category variation may affect attractor strengths as well as shifts in their definitions, which have been shown to vary as a function of semantics, giving different 'prototypes' (Gili Fivela, 2013). Their model thus creates a phonetic space that integrates phonetics and phonology without giving up the distinction between them, in accordance with the call for a hybrid model by Pierrehumbert (2016), i.e., a model which captures an abstract level along with a richly detailed level of representation of the phonetic dynamics of indexical features, regular sound changes, and word-specific phonetics. An understanding of the functions of intonation must therefore be incomplete if we concentrate our efforts on discrete contrasts. This critical point was made in somewhat graphic terms by Dwight Bolinger (1978, p. 475) as follows: "Intonation is a half-tamed savage. To understand the tamed or linguistically harnessed half of him one has to make friends with the wild half." We intend to do this by evaluating the lexical and intonational contrasts of Zhumadian Mandarin in a categorical perception experiment.

Section 1.1 of this 'Introduction' briefly sketches the state of the evidence for discrete or gradient representations of pitch contours; Section 1.2 introduces the two language groups involved in our study; and Section 1.3 examines the research paradigm we used.

1.1. EVIDENCE OF DISCRETENESS IN PITCH CONTOURS

The categorical status of lexical tone contrasts is uncontroversial in most of the linguistic literature (e.g., Bao, 1999; Best, 2019; Hyman, 2011; Yip, 2002), and evidence for it is a frequent by-product of investigations addressing different research questions. Neuroscientific evidence for the phonological nature of lexical tone contrasts has come from ERP studies. Larger mismatch negativities (MMNs), associated with deviant processing, have been found in the left hemisphere across Mandarin Chinese tone categories than within them, which has been advanced as support for the categorical nature of Chinese tone contrasts (Shen & Froud, 2019; Xi, Zhang, Shu, Zhang, & Li, 2010). Similarly, Zheng, Minett, Peng, and Wang (2012) report categorical effects in behavioural identification results for across-category rising vs. level pitch contours from Mandarin and Cantonese listeners, with the latter group additionally showing a larger

P300 effect, commonly associated with stimulus categorization. Neuroscientific evidence has also come from data showing activity in the right hemisphere prior to increased activity in the left, indicating a transfer from phonetic processing to phonological processing (Zhang, Xi, Xu, Shu, Wang, & Li 2011).

The picture for intonation is more varied. European language groups like Swedish/Norwegian and Dutch/German include varieties with binary lexical tone contrasts. Thus, in Cologne German, *Schav* /ʃaaf/ spoken with an early pitch fall is the word for ‘slicer’, while *Schaaf* /ʃaaf/ spoken with a late fall is ‘cupboard’, for the same declarative intonation (e.g., Gussenhoven & Peters 2004). In an oddball experiment in which either a lexical tone or an intonational melody were used as deviants, speakers of the related tonal dialect of Roermond displayed predominant activation of the left temporal cortex for tonal, but not for intonational, contrasts. However, a non-tonal Dutch control group, for whom all contrasts were intonational, showed more general left-hemisphere activation for both types of contrast (Fournier, Gussenhoven, Jensen, & Hagoort, 2010), in line with similar findings for Japanese (Imaizumi, Mori, Kiritani, Hosoi, & Tonoike, 1998). This may seem to suggest that discreteness of intonation depends on the presence of tone in the same system. Another interaction of this sort was reported by Liang and van Heuven (2007), who found that speakers of Uygur, an intonation-only language, processed intonation contrasts more efficiently than speakers of tonal Sinitic languages. A non-ambiguous result was obtained by Chien, Friederici, Hartwigsen, and Sammler, (2020), who administered three identification tasks during MRI recordings to Standard Chinese and German speakers using morphed continua between Standard Chinese monosyllables, a statement–question intonation contrast, a tone contrast (rise vs. fall), and a gender contrast. The groups performed similarly on the intonation and gender identification tasks, with gender yielding sigmoid response curves and intonation approximately linear ones, but for the tone contrast the Chinese group provided responses similar to those for gender, and the German group similar to those for intonation. Their fMRI data were in conformity with these results, indicating that intonation was gradient for both groups. While neuroscientific evidence for intonational categoriality is thus scant, behavioural evidence has been obtained from a variety of experimental approaches, such as an imitation task (Pierrehumbert & Steele, 1989), semantic judgements (Borràs-Comes, Vanrell, & Prieto, 2014; Gussenhoven & Rietveld, 2000; Vanrell, Mascaró, Torres-Tamarit, & Prieto, 2012), and equivalence judgements (Odé, 2005). Arguably, these results may not yet have been sufficiently replicated across intonational contrasts and languages, so they may retain a provisional character (Gussenhoven, 2006; Prieto, 2012).

Our contribution to this debate is threefold. First, we intend to show how four phonetically different falling pitch shapes representing two lexical tones

spoken with statement and question intonations in Zhumadian Mandarin are perceived as categorically different when signalling the lexical contrast, but as gradually different when signalling the intonational (statement vs. question) contrast. We equally intend to do this for four rising pitch contours. Second, we demonstrate that the classical Categorical Perception method can be used for this purpose, provided it includes a baseline for the stimuli as well as a baseline for the language whose contrasts are under investigation. Third, we use a novel approach for the statistical analysis, where we model categorical perception curves directly using polynomials.

1.2. THE LANGUAGES IN THE EXPERIMENT

Our experiment involved two groups, a control group of Indonesian listeners and an experimental group of Zhumadian Mandarin listeners.

Indonesian has neither lexical tone nor, *pace* a number of earlier claims, does it have stress. Evidence of stresslessness is found in the absence of perceived prominence by native speakers for phrase-final rising–falling pitch movements (Odé, 1994), and in the absence of a consistent syllabic alignment of the rising–falling movement in varieties of Malay (Goedemans & van Zanten, 2007; Maskikit-Essed & Gussenhoven, 2016). Also, Indonesian listeners differ from Dutch listeners, but not from French listeners, in their performance on the Sequence Recall Task (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Peperkamp, Vendelin, & Dupoux, 2010), indicating that they are ‘stress deaf’ in the way French listeners are (Rahmani, Rietveld, & Gussenhoven, 2015). Pitch range differences have never been claimed to be categorical in the language. Indonesian does have a contrast between rising f_0 , as used to mark pre-final phrases as well as final phrases in question sentences, and (rising–)falling pitch, marking final declarative phrases (Halim, 1974; Stoel, 2006). Since rising vs. falling pitch contrasts play no role in our experiment, the strong prediction is that Indonesian listeners perceive the pitch alignment and pitch range differences involved in the Zhumadian pitch contrasts non-categorically.

The *Mandarin dialect of Zhumadian* (Henan Province, China)¹ contrasts four monosyllabic lexical tones, two rises and two falls, one with early and one with late alignment. The two early aligning tones begin the pitch movement more or less at the beginning of the vowel, the two late-aligning tones begin it halfway through the syllable rhyme. The tones are here referred to as ‘Early

¹ The city of Zhumadian, the capital of the eponymous prefecture in Henan Province (PRC), has over 500,000 inhabitants. Many older inhabitants are monolingual speakers of the dialect with a receptive knowledge of Standard Mandarin, while younger speakers are bilingual with Standard Mandarin. In primary and secondary education, students are asked to speak Standard Mandarin during class, but outside class and outside school the dialect is used exclusively.

Fall (EF)', 'Late Fall (LF)', 'Early Rise (ER)' and 'Late Rise (LR)'. Question versions of these tones are pronounced with raised or expanded pitch range. That is, question intonation does not alter the direction of the pitch movement relative to their declarative versions. While otherwise close to Standard Mandarin, its specific alignment contrasts in its lexical tone system provided a rare opportunity to test the status of lexical and intonational contrasts for categoriality *within sets of uni-directional contour shapes*, a unique feature of our experiments. Lexical alignment contrasts are rare, and were earlier believed to be universally excluded, as noted by Remijsen (2013). For falls, they have been reported for Dinka by Remijsen (2013) and for Shilluk by Remijsen and Ayoker (2014), while alignment contrasts for rises have been reported by DiCanio, Amith, and Castillo García (2014) for YoloXóchitl Mixtec. There is no consensus on the categorical (or gradient) status of Mandarin intonation. The contrast between statement and question intonation has been described in terms of boundary tones (Chao, 1968; Duanmu, 2000), as well as in terms of register raising and range expansion (Liu & Xu, 2005; Shen, 1990; Xu, 2015; Zhang, 2018). This raised pitch may be due to a final boundary tone but might equally be assumed to be a paralinguistic feature.

1.3. CATEGORICAL PERCEPTION

In order to evaluate the hypothesis that the Zhumadian Mandarin contrasts between early and late falling lexical tones and between early and late rising lexical tones are perceived categorically by native listeners, while their statement and question pronunciations are perceived as gradiently different, we used a classical categorical-perception experiment. It involves two perception tasks with one or more sets of stimuli representing acoustic continua between two phonetic forms (Eimas, 1963; Liberman, Harris, Hoffmann, & Griffith, 1957). A discrimination task is used to establish to what extent adjacent stimuli in the continuum are heard as the same or different, while an identification task is used to assess the degree to which stimuli are heard as either one (putative) category or the other. A categorical-perception effect is reflected in a greater sensitivity to the phonetic difference between adjacent stimuli at or bordering on the midpoint of the continuum than to similarly spaced stimulus pairs within each half of the continuum (discrimination task) as well as by a sigmoid response curve crossing the midpoint, indicating a relatively sudden perceptual shift from one form to the other (identification task).

The earliest categorical perception experiment involved place contrasts for English prevocalic voiced plosives, which are strongly cued by the transition of the second formant into the vowel (Liberman et al., 1957). A 14-step continuum in which the second-formant transition into [e] was varied from falling to rising yielded sharp perceptual shifts from [be] to [de] and from [de] to [ge] in

an identification task (these results were replicated and refined by Hall & Peck, 2017). Additionally, discrimination performance was best at the perceptual shifts from one plosive to the next for pairs of stimuli that were one and two steps apart. Similarly, regular effects were found by Abrahamson and Lisker (1970) for the prevocalic voicing contrast in plosives, as in English *bear* vs. *pear*. However, experiments with vowel quality have yielded less distinct results, which has been attributed to the inherently more gradient nature of the continua (Fry, Abramson, Eimas, & Liberman, 1962; Gerrits & Schouten, 2004; Liberman, Mattingly, & Turvey, 1972, cf. also Repp, 1984). Moreover, because categorical effects have been obtained with non-speech stimuli (e.g., Cutting, Rosner & Foard, 1976), at least part of the effect is due to acoustic properties of the contrast. For voice pitch, Kohler (1987) and Schneider and Lintfert (2003) successfully demonstrated the categorical nature of, respectively, early versus late f_0 peaks and final low versus rising pitch in German intonation, but overall, intonation contrasts have failed to produce consistently categorical results (Post 2000; Remijsen & van Heuven, 1999). The method has come under attack for two further reasons. First, the relation between the identification and discrimination results has varied across experiments, implying that there is no criterial definition of the degree of agreement between the results of the two tasks (Gerrits & Schouten, 2004). This problem interacts with that of the acoustic nature of the stimuli, because vowel and pitch continua may show sharp transitions in the identification results without concomitant sensitivity increases around that boundary in the discrimination data (Savino & Grice 2011). Second, Gerrits and Schouten (2004) found that only one of two tasks, one that compared different sizes of a difference along a phonetic dimension, yielded the categorical-perception effect in an [i ~ u] vowel contrast, implying there may be crucial task effects.

Importantly, the above disadvantages of the categorical-perception method can be avoided. First, since it is not a binary diagnostic test, it cannot be used to determine the nature of a single phonetic contrast. It can, however, be used to establish significant differences between multiple such contrasts that include a baseline. Minimally, therefore, there should be an uncontroversially gradient (or categorical) contrast and a hypothetically categorical (or gradient) one, so that results can be presented in terms of statistical differences between the experimental and the baseline stimuli. Second, the stimulus sets must be acoustically homogeneous and so avoid comparisons between acoustically disparate contrasts. Moreover, since there may still be a risk that specific acoustic differences in one or more of the acoustic continua are prone to categorical perception as a result of some quantal feature (Stevens, 1989), a control group should be used whose language does not employ either kind of contrast in its grammar. Any categorical effect in the experimental group will thus only be attributable to the discreteness of the contrast if it fails to appear in the control

group. Equally, the absence of a categorical effect in the experimental group will only be acceptable as a meaningful result if it equally fails to appear in the control group. In establishing statistically significant differences, we advance the methodology in this domain by assessing categorical perception with models that include polynomial terms. Specifically, we model discrimination curves with quadratic terms and identification curves with cubic terms. Evidence for categorical perception should emerge in the form of significant cubic effects for identification and significant quadratic effects for discrimination.

2. Materials

All continua were created from recordings made by a female native speaker of Zhumadian Mandarin in a studio at the Department of Psychology of Penn State University. She pronounced isolated statement and question pronunciations of the three quadruplet tonal minimal pairs in Table 2 immediately after saying a brief statement or question containing the word in final position. Words were presented in a random order on separate slides on a laptop. We selected one statement and one question source utterance from minimally three recorded utterances, avoiding ones with creak, which was not always possible in the case of the declarative Late Rise and Late Fall. We used the STRAIGHT morphing synthesis procedure (Kawahara, Masuda-Katsuse, & Cheveigné, 1999) to create seven stimuli for each of the dimensions depicted as the sides of the two squares in Figure 1, following Escudero, Benders, and Lipski (2009). Horizontal continua are between lexical tones, and vertical ones between intonation contours, with falling and rising tones appearing in separate quadrants. The dots represent 28 stimuli in each quadrant, since the corners represent extreme stimuli on two continua. All continua turned out to yield natural-sounding stimuli. The total number was 7 (steps) \times 4 (continua) \times 3 (segmental syllables) \times 2 (contour shapes) = 168.

Durations of the 24 selected source utterances are given in Table 3. On average, words in questions are 27% longer than in statements, while the duration of the sonorant onsets ([l] and [m]) tends to be longer as their rhymes are shorter (Pearson's $r = -0.64$). Pitch contours of the recorded words that

TABLE 2 *Four tonal minimal quadruplets used in the experiment*

	[dii]		[lyy]		[mae]	
Early Fall	'earth'	地	'to filter'	滤	'sell'	卖
Late Fall	'opponent'	敌	'donkey'	驴	'bury'	埋
Early Rise	'bottom'	底	'aluminum'	铝	'buy'	买
Late Rise	'low'	低	'green'	绿	'cereal'	麦

CATEGORICAL TONE AND GRADIENT INTONATION

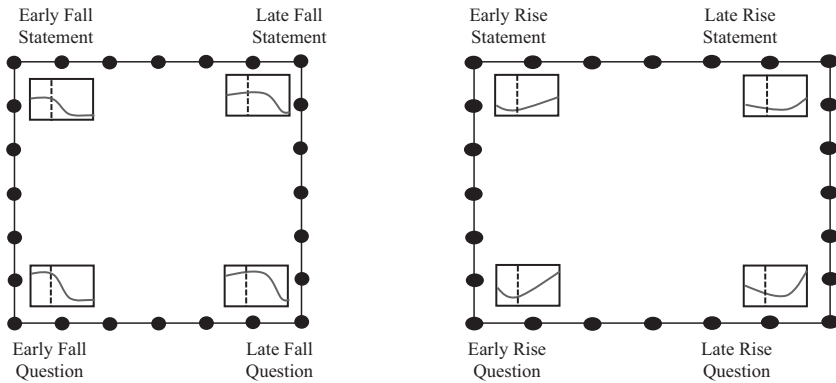


Fig. 1: Acoustic continua between Early Fall and Late Fall and Early Rise and Late Rise in statement and question intonations and between their respective intonations, with schematic pitch contours, including an indication of the CV boundary by a dotted line.

were used as the source utterances for the continua are shown in Figure 2, with declarative intonations on the left. The syllable [lyy] tends to differentiate Early Fall and Late Fall on the basis of peak alignment rather than fall alignment of the sort shown for [dee] and [mae].

3. The control group: Indonesian participants

A pitch identification task and a pitch discrimination task were implemented as multiple-forced choice experiments in Praat (Boersma & Weenink, 2017). They were administered to 45 self-declared native speakers of Indonesian (Bahasa) from the university student population in Malang (35 female; mean age 22.2 years, with a range of 18 to 25) in a classroom in which four laptops were used simultaneously, each attended by an experimenter. Subjects listened to the stimuli via headphones or, whenever they felt that these fitted more comfortably underneath their headscarves, their personal earphones. The experiments were self-paced, whereby the pitch identification experiment preceded the discrimination experiment. Participants were told that they could abandon the session at any time without providing an explanation. All participants agreed to take part in the second experiment. They were paid a 10-euro fee.

3.1. THE INDONESIAN PITCH IDENTIFICATION EXPERIMENT

In the pitch identification experiment, stimuli were blocked by the eight continua in Figure 1. Each of the three syllables was minimally represented

TABLE 3. Durations of onsets (C) and rhymes (V) of the selected statement and question pronunciations of the words in Table 2. In the case of [dii], the duration is the VOT, the closure phase of [d] being voiceless.

Contour	[dii]		[lyy]		[mae]		Mean	
	C	V	C	V	C	V	C	V
Early Fall_S	9	208	108	261	130	215	119	238
Late Fall_S	14	215	117	194	75	315	96	255
Early Fall_Q	8	380	91	395	118	287	105	341
Late Fall_Q	8	381	77	286	91	388	84	337
Early Rise_S	31	283	93	334	162	125	128	230
Late Rise_S	14	281	63	262	125	191	94	227
Early Rise_Q	17	441	76	469	142	278	109	374
Late Rise_Q	14	329	85	310	131	300	108	305
Mean S	17	247	95	263	123	212	109	237
Mean Q	12	383	82	365	121	313	101	339

in one and maximally in three continua, but assignment of syllables to continua was otherwise random. For the tone continua, we used the stimuli from the statement continua for lyy^{EF-LF} and dee^{ER-LR} and from the question continua for dee^{EF-LF} and lyy^{ER-LR} ; for the intonation continua, we used the stimuli for dee^{LF} , mae^{EF} , mae^{ER} , and mae^{LR} . Each block was preceded by a practice session with an exemplar from one end of the scale which participants were instructed to associate with a button labelled A, followed by an exemplar from the other end which they were instructed to associate with a button labelled B. This procedure was repeated with two exemplars representing the second and sixth steps. Each of the eight experimental blocks contained two copies of each of the seven steps in the continuum. The fourteen trials in each block were randomized per participant. Two instances of the stimulus were presented at each trial after a 650 ms warning sound and a pause of 350 ms, with a 300 ms pause before the repetition. The order of the eight blocks was reversed for 22 of the 45 participants. Due to a technical error, the practice session for the statement-to-question continuum of mae^{ER} was based on incorrect exemplars, for which reason those scores were discarded. The analysis is thus based on 7 (continua) \times 7 (steps) \times 2 (repetitions) \times 45 (participants) = 4,410 scores.

3.1.1. Results of the identification experiment: Indonesian participants

All data, analysis code, and stimulus materials are accessible in the following Open Science Framework repository: <https://osf.io/79h5z/?view_only=2ba135ab79c3409c9d8b7e499e7aa7ce>. Since identification experiments typically yield sigmoid response curves for categorical contrasts, we applied polynomial logistic mixed-effects regression analyses with the logit link

CATEGORICAL TONE AND GRADIENT INTONATION

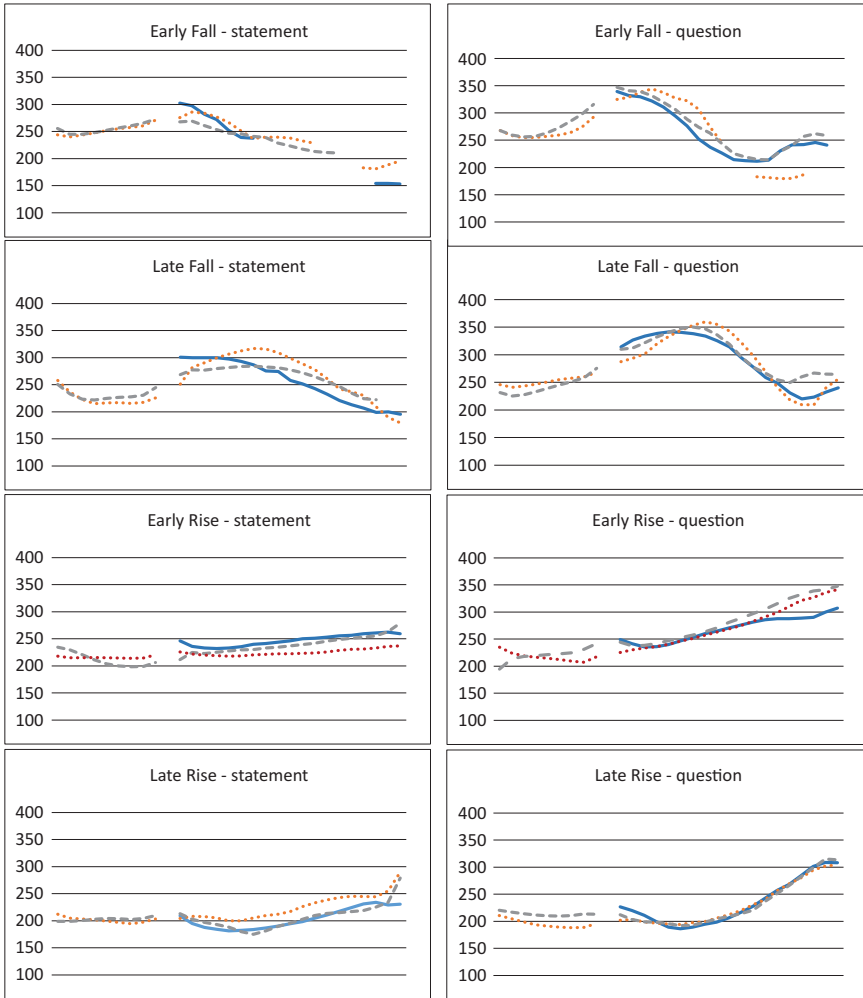


Fig. 2: F0 tracks over 19 measuring points for C and V separately of the selected statement and question pronunciations of the words in Table 2 which were used as the source utterances for the continua. Normalized time, with C measured in [l] and [m] and depicted over a stretch of 45% of that of V, approximating the average C/V duration ratio. — = [dii], = [lyy], - - - - = [mae].

function (see, e.g., Jaeger, 2008) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Development Core Team, 2015). We included the dependent variable *response* (early/late for tone identification and statement/question for intonation identification; question/late responses were coded as '0' and statement/early responses as '1') and the independent variables

gender, *syllable* ([dii], [lyy], or [mae]), *step* (on an auditory continuum; 1–76), *type of contour* (fall/rise), and *type of contrast* (tone/intonation), with treatment coding for factors. In addition, we included a random intercept for *listener*, as means or ranges of identification scores by individual listeners may vary. We obtained the fixed-effects structure of the model by means of stepwise variable selection, in which we included variables if they significantly improved the model fit, as established by means of likelihood ratio tests. Importantly, whenever a variable that contained quadratic or cubic terms (e.g., *step*) was included in an interaction (e.g., between *step* (*linear*) and *type of contrast*), we included similar interaction terms for the quadratic and cubic terms (in this case between *step* (*quadratic/cubic*) and *type of contrast*), for the sake of interpretability. Similarly, whenever a quadratic or cubic term of a variable was included in the analysis for Zhumadian, this quadratic or cubic term was also included in the analysis for the Indonesian counterpart and vice versa, for the sake of comparability. Continuous variables were standardized to a mean of zero and centred (Belsley, Kuh, & Welsch, 1980). After determining the fixed effects structure, we tested, for each fixed effect, whether the inclusion of a random slope improved the model fit (trying random slopes for all significant fixed effects), by means of likelihood ratio tests (Baayen, 2008). This procedure for model selection, standardization of continuous variables, and testing for significance was employed for all logistic mixed-effects regression analyses reported in this study. Unless noted otherwise, the *p*-values of the logistic mixed-effects regression analyses are based on Wald's *z* statistic, while the beta coefficients are logit coefficients.

Table 4 shows main effects of *type of contour* and *type of contrast*. Participants provided fewer Early/Statement responses for rising than for falling contours and more Early/Statement responses for tone contrasts than for intonation contrasts. The quadratic effect, *step* (*quadratic*), is due to the negative dents in the response curves more or less at the perceptual halfway points of the continua. That is, proximity to the continuum endpoints increases the identifiability of stimuli in the left half of continua to a greater extent compared to those in the right half of continua. The random slopes of *step* (*linear*) and *type of contrast* for *listener* indicate that the identification curves along the auditory continua, as well as the differences between tone and intonation identification, varied significantly across participants. The results for *type of contour*, *type of contrast*, *step* (*linear*), and *step* (*quadratic*) are shown graphically in Figure 3.

Subsequently, we analyzed participants' reaction times (RTs) by means of linear mixed-effects models with treatment coding for factors. We included the dependent variable *RT* (logged RT) and the same fixed and random variables as for the logistic analyses reported above. For all the RT analyses reported in this study, we only included data with standardized residuals

TABLE 4. Results of the logistic mixed-effects regression for the Indonesian identification experiment (0 = Late/Question, 1 = Early/Statement)

Random effects		Standard Deviation		
Listener		0.46		
Listener: Step (linear)		0.93		
Listener : Type of contour (rise)		0.76		
Type of contrast (tone)		0.71		
Fixed effects	Estimate	<i>z</i>	<i>p</i>	
Intercept	-0.320	-3.26	< .01	
Type of contour (rise)	-0.340	-2.52	< .05	
Step (linear)	-1.045	-6.25	< .0001	
Step (quadratic)	0.151	3.59	< .001	
Step (cubic)	0.092	1.79	= .07	
Type of contrast (tone)	0.379	2.95	< .01	

between -2.5 and 2.5, in line with Baayen (2008). We then re-ran the mixed-effects regression models.

Table 5 shows a main effect of *syllable*, indicating slower RTs for words with [lyy] and [mae] than those with [dii], the only syllable that agrees with Indonesian phonology, which probably explains the faster processing by these participants of stimuli with that syllable. An interaction between *type of contour* and *type of contrast* indicated that falling stimuli in the intonation continua were processed faster than rises and falls in the tone continua, possibly due to their similarity to falling pitch in the stimuli in the intonation continua. There was also a main effect of the quadratic term of *step*, which showed somewhat faster RTs near the extremes of the continua.

3.2. THE INDONESIAN PITCH DISCRIMINATION EXPERIMENT

We selected stimulus pairs from one of each of the eight continua in Figure 2 that were 1, 2, 3, and 4 steps apart, which yielded 18 stimulus pairs for each continuum, a total of 144 stimulus pairs. For the tone continua with statement carrier intonation, we used lyy^{EF-LF} and dii^{ER-LR} , while for the tone continua with question carrier intonation we used mae^{ER-LR} and lyy^{EF-LF} , and for the intonation continua we used lyy^{EF} , lyy^{LF} , mae^{ER} , and dii^{LR} . Two fillers were added that were 6 steps apart. A training session with a pair of identical stimuli, a pair with stimuli 6 steps apart, and a pair 5 steps apart preceded the presentation of all 146 stimulus pairs in a random order to each participant. Each stimulus pair came after a warning sound of 650 ms followed by 350 ms of

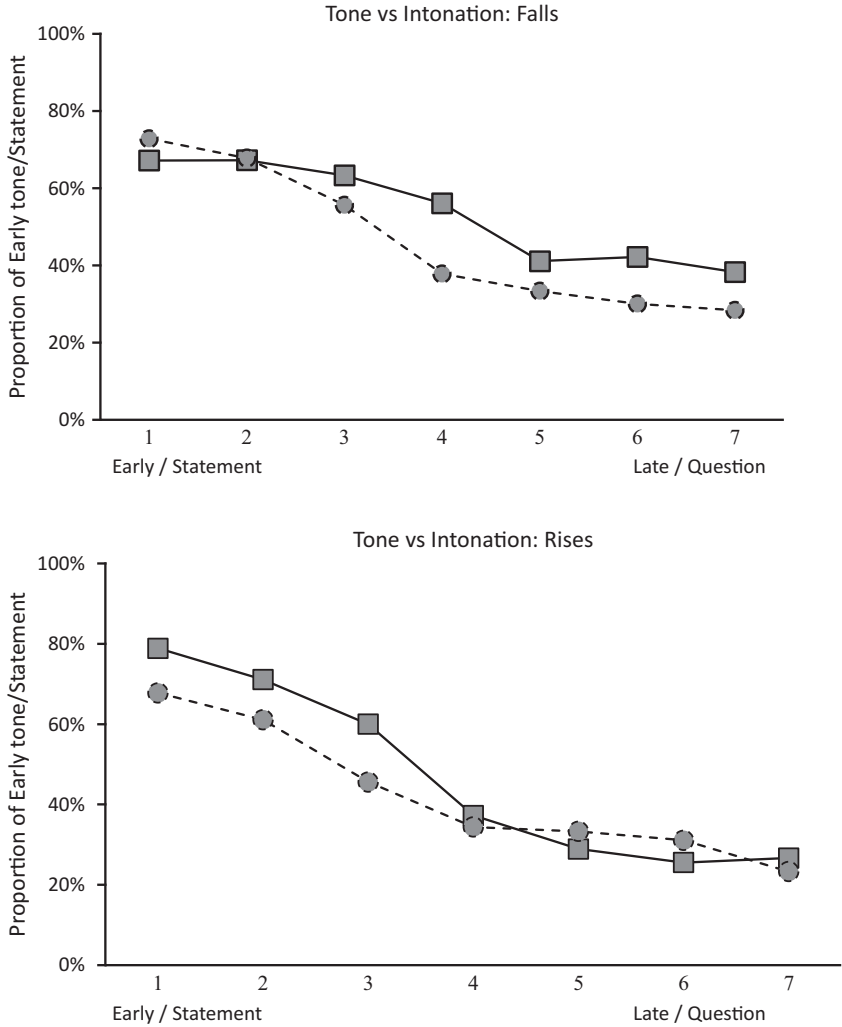


Fig. 3: Identification response curves for falls (top) and rises (bottom) by step for stimuli from intonation continua (Statement vs. Question: ---●---) and tone continua (Early vs. Late: —■—) by Indonesian participants, with weak negative dents at around steps 3 or 4.

silence and a 400 ms interval between the first and the second stimulus. A ‘Replay’ button allowed participants to listen to the same pair as often as they wished. They were instructed to give their judgement by clicking either a button labelled *sama* ‘same’ or one labelled *berbeda* ‘different’.

TABLE 5. Results of the linear mixed-effects regression analysis of the logged RTs for the Indonesian identification experiment

Random effects	Standard Deviation	
Listener	0.13	
Listener : Type of contour (rise)	0.05	
Listener : Syllable (lyy)	0.09	
Listener : Syllable (mae)	0.09	
Listener : Type of contrast (tone)	0.10	
Fixed effects	Estimate	t
Intercept	1.532	75.51
Type of contour (rise)	0.057	5.22
Syllable (lyy)	0.032	2.24
Syllable (mae)	0.089	5.68
Step (linear)	0.002	0.26
Step (quadratic)	-0.007	-2.69
Step (cubic)	0.001	0.35
Type of contrast (tone)	0.073	4.16
Type of contour (rise) : type of contrast (tone)	-0.087	-8.34

3.2.1. Results of the pitch discrimination experiment: Indonesian participants
 Since discrimination experiments typically yield a non-linear discrimination function for categorical contrasts, with more ‘different’ responses near the perceptual category boundary, we applied the same analysis on the (45 × 144 =) 6,480 scores as for the identification experiment, including the dependent variable *response* (Same/Different) and the same independent variables as for the identification experiment, in addition to *step difference* (step difference between the two stimuli in a trial).

Table 6 shows main effects of *step difference* and *syllable*. As expected, participants consistently provided more ‘different’ responses for word pairs with larger step differences, as shown in Figure 4, where response functions never cross. There were also more ‘different’ responses for stimulus pairs that contained the syllable [mae] and fewer for pairs that contained the syllable [lyy] compared to those that contained [dii]. Both the *step difference* and the *syllable* effects varied significantly across participants, as shown by the random slopes of *step difference* and *syllable* for *listener*. In addition, we again found two-way interactions between *type of contrast* and *step* and between *type of contrast* and the quadratic term of *step*, ‘*step (quadratic)*’. To interpret these interactions, we split the data by *type of contrast*. For the tone discrimination data, we found a significant effect of *step* ($\beta = 0.107, z = 2.00, p < .05$), but no effect of the quadratic term of *step* ($\beta = 0.023, z = 0.47, p = .64$), indicating that the

TABLE 6. Results of the logistic mixed-effects regression analysis for the Indonesian pitch discrimination experiment (0 = Same, 1 = Different)

Random effects	Standard Deviation		
Listener		0.86	
Listener : Step difference		0.36	
Listener : Syllable (lyy)		0.72	
Listener : Syllable (mae)		1.04	
Fixed effects	Estimate	z	p
Intercept	-1.290	-8.47	< .0001
Step difference	1.158	17.30	< .0001
Syllable (lyy)	-0.812	-5.86	< .0001
Syllable (mae)	1.253	7.09	< .0001
Step (linear)	-0.334	-6.76	< .0001
Step (quadratic)	0.312	6.66	< .0001
Type of contrast (tone)	0.121	1.28	= .20
Step (linear) : type of contrast (tone)	-0.307	-4.69	< .0001
Step (quadratic) : type of contrast (tone)	0.414	5.99	< .0001

Indonesian participants show a heightened sensitivity to differences closer to the right end of the tone continuum, which trend is discernible in the top panel of Figure 4. For the intonation discrimination data, we found significant effects of *step* ($\beta = -0.346$, $z = -6.95$, $p < .0001$) and *step (quadratic)* ($\beta = 0.291$, $z = 6.14$, $p < .0001$). The step effect is discernible in the heightened sensitivity to stimulus pairs located at the left edge of the continua, indicating a somewhat better performance for statement continua than question continua (bottom panel, Figure 4). More impressively, Indonesian participants show a reduced sensitivity to phonetic differences closer to the midpoint of the intonation continuum.

Subsequently, we analyzed participants' RTs by means of linear mixed-effects, including the same dependent and independent variables as for the pitch identification experiment.

Most relevantly, we found main effects of *step difference* and the quadratic term of *step difference*, indicating that RTs were slower for word pairs with larger step differences (Table 7). This shows that smaller step differences were rapidly judged to be the same by many subjects, while larger step differences were more difficult to judge, but with the largest step difference no longer increasing the difficulty. The main effect for *tone* indicates that the intonation pairs were slightly more slowly discriminated than the tone pairs, which effect was stronger at the right end of the continua, as indicated by the interaction between *type of contrast* \times *step (quadratic)*. The main effect of *syllable* is due to slower responses to word pairs containing the syllables [mae] and [lyy] than to those containing [dii].

CATEGORICAL TONE AND GRADIENT INTONATION

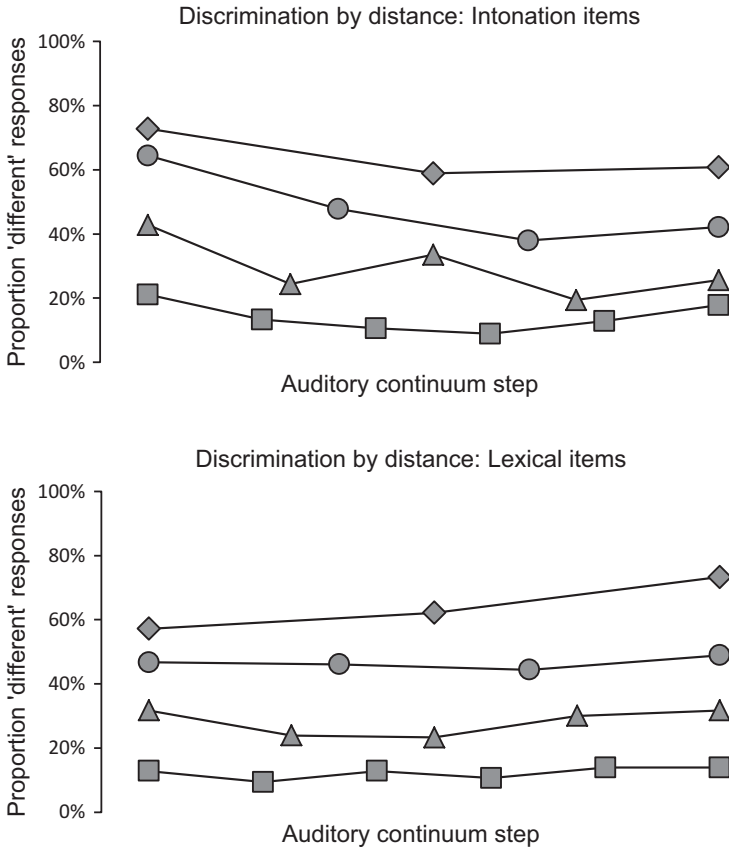


Fig 4: Discrimination results for Indonesian participants for intonation contrasts (top) and lexical contrasts (bottom) for 1 (■), 2 (▲), 3 (●), and 4 (◆)-step distances between members of pairs of stimuli.

3.3. DISCUSSION OF THE RESULTS OF THE INDONESIAN PARTICIPANTS

The results of the identification and discrimination experiments support the hypothesis that the intonation and tone continua represent gradient scales for the Indonesian participants. Neither continuum type yielded a sigmoid response curve in the identification experiment, which would have shown up as a significant cubic effect. Neither did either type of continuum yield dome-shaped response curves in the discrimination test, which would have shown up as a positive effect of *step* (*quadratic*). Reverse effects (i.e., negative *step* (*quadratic*) effects) were found instead. The interaction with *type of contrast* indicates that discriminability increased towards the extremes of the

TABLE 7. *Results of the linear mixed-effects regression analysis of the logged RTs for the Indonesian discrimination experiment*

Random effects	Standard Deviation	
Listener	0.07	
Fixed effects	<i>Estimate</i>	<i>t</i>
Intercept	1.619	149.02
Step difference (linear)	0.012	8.46
Step difference (quadratic)	-0.006	-3.78
Syllable (lyy)	0.038	12.56
Syllable (mae)	0.051	16.09
Step (linear)	0.004	2.65
Step (quadratic)	0.003	1.73
Type of contrast (tone)	-0.004	-1.07
Type of contrast (tone) : Step (quadratic)	-0.005	-2.33

intonation continua, but that for the tone continua discriminability increased towards the ‘early’ end only. This indicates a perceptual diffuseness for the Indonesian listeners of later alignments. Further effects, again, have no bearing on our hypothesis, but suggest that familiar stimuli are processed better and faster. In our case, this applied to falling intonation in final position and for the syllable [dii], as revealed by the results of the two Response Time analyses.

4. The experimental group: Zhumadian participants

The identification experiment and the discrimination experiment were similarly implemented as multiple-forced choice experiments in Praat and presented to 40 participants (29 female; mean age 18.9, with a range of 17 to 21), recruited from Huanghuai College in central Zhumadian (22 participants) and Queshan Senior Middle School (18 participants) in Queshan, on the city’s southern edge. All of them reported to be native speakers of the Zhumadian dialect; 25 of them declared that both their caregivers were local native speakers. Of the remaining 15 participants, 14 declared that one caregiver came from a different location. One participant, who was judged to be a native speaker by three other participants, declared that both caregivers originally came from a different location. All participants declared familiarity with the 12 words in Table 2. In each location, the experiments were presented on laptops placed on desks in different parts of a classroom. The experiments were self-paced and were done during consecutive 30-minute slots, whereby the identification experiment preceded the discrimination experiment. Like the Indonesian participants, the Zhumadian participants were remunerated with a 50-CNY fee and advised on their freedom to abandon the experiment.

4.1. THE ZHUMADIAN TONE AND INTONATION IDENTIFICATION EXPERIMENT

For the tone and intonation identification experiment, the 168 stimuli were arranged in six blocks, each block containing the 28 stimuli for the four continua constructed from the two words in each of the six double-lined pairs of cells in Table 2, two tone continua between the two words and two intonation continua between statement and question versions of each word (cf. - Figure 2). Stimuli were randomized per block and presented twice after a 650 ms warning sound and a pause of 350 ms, with a 400 ms pause before the repetition. Participants were asked to decide which of the two words they heard by clicking one of two boxes containing the character for the two words. On the same screen beneath these boxes, two further boxes were shown, one with the Chinese full stop and one with a question mark, which they were instructed to choose between depending on whether they thought the word was spoken as a statement or a question. Before each block, four exemplars of the relevant words were presented with feedback in each case.

Step 2 of the [dii] continua with rises was missing (4×40 responses), so that we obtained 6,560 scores instead of the intended $40 \times 168 = 6,720$. Moreover, if a participant identified a stimulus from a statement tone continuum as a question pronunciation or vice versa, we assumed that the recognition of the word was based on a false assumption of the carrier intonation. Similarly, if a stimulus from a late-alignment intonation continuum was identified as a word with early tone alignment or vice versa, we assumed that the recognition of the intonation was based on a false assumption of the carrier word. By this criterion, 617 scores were discarded (9.4%). As for the misidentification of carrier tones, shown in the left-hand panel of Figure 5, Early Rises were less often misidentified than the three other tones. Incorrect identifications of the carrier intonation were considerably more frequent (6.66%) than incorrect identifications of the carrier word (2.74%). The right-hand panel shows a bias towards statement intonation in the misidentifications of the carrier intonation. Most of these occurred in the continua for rising tones, which indicates that question intonation is more easily mistaken for declarative intonation than the other way around, and more so on rising tones. In Section 4.2.1 we analyse the remaining 5,943 valid scores.

4.1.1. *Results of the tone and intonation identification experiment: Zhumadian participants*

We again applied polynomial logistic mixed-effects regression analyses with the dependent variable *response* and the same independent variables as for the

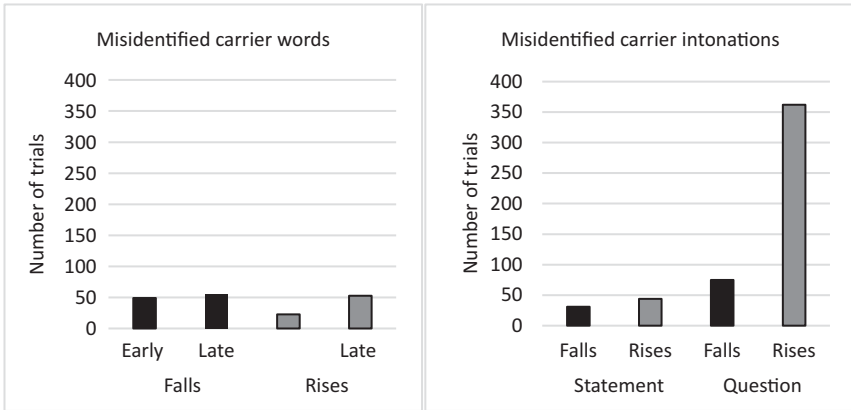


Fig. 5: Misidentified carrier tones from intonation continua for Early Falls, Late Falls, Early Rises, and Late Rises (left panel), and misidentified carrier intonations of stimuli from rising and falling tone continua with question intonation and statement intonation (right panel) by Zhumadian participants. Black bars = falls, grey bars = rises.

analyses of the Indonesian identification experiment. The results are provided in Table 8.

The main effect of *type of contour* indicates that participants provided more Early/Statement responses for falling than for rising contours. The main effect for *syllable* indicates more Early/Statement responses for words with the syllable [dii]. The differences between falling and rising contours varied significantly across participants, as shown by the random slope of *type of contour* for *listener*. More importantly, we found main effects of *step* and the quadratic term of *step* as well as two-way interactions between *step* and *type of contrast* and between the cubic term of *step* and *type of contrast*. The identification curves along the auditory continua, as well as the differences between tone and intonation identification, varied significantly across participants, as shown by the random slopes of *step* and *type of contrast* for *listener*. In order to interpret the two-way interactions, we split the data by *type of contrast*.

For the tone identification data, we found significant main effects of *step* ($\beta = -3.108$, $z = -17.78$, $p < .0001$), *step (quadratic)* ($\beta = 0.344$, $z = 4.13$, $p < .0001$), and *step (cubic)* ($\beta = 0.453$, $z = 4.59$, $p < .0001$). For the intonation identification data, we found significant main effects of *step* ($\beta = -1.551$, $z = -11.43$, $p < .0001$) and *step (quadratic)* ($\beta = 0.325$, $z = 5.01$, $p < .0001$), but no effect of *step (cubic)* ($\beta = 0.0004$, $z = 0.01$, $p = 1.00$). The resulting tone and intonation identification curves are given in Figure 6, where the tone identification scores yield a pronounced sigmoid response curve, while the intonation identification scores do not. The weaker interaction between *step* and *type of contrast* is

CATEGORICAL TONE AND GRADIENT INTONATION

TABLE 8. Results of the logistic mixed-effects regression analysis for the Zhumadian identification experiment (0 = Late/Question, 1 = Early/Statement)

Random effects	Standard Deviation		
Listener	0.60		
Listener : Step (linear)	0.32		
Listener : Type of contour (rise)	0.43		
Listener : Type of contrast (tone)	0.50		
Fixed effects	Estimate	z	p
Intercept	0.680	5.19	< .0001
Type of contour (rise)	-0.303	-3.09	< .01
Syllable (lyy)	-0.465	-5.36	< .0001
Syllable (mae)	-0.484	-5.63	< .0001
Step (linear)	-1.517	-11.54	< .0001
Step (quadratic)	0.322	5.04	< .0001
Step (cubic)	0.003	0.05	= .96
Type of contrast (tone)	-0.571	-4.63	< .0001
Step (linear) : type of contrast (tone)	-1.625	-7.88	< .0001
Step (quadratic) : type of contrast (tone)	0.003	0.03	= .98
Step (cubic) : type of contrast (tone)	0.462	3.83	< .001

caused by the bias away from question identifications, which appears in the continua both for falling and for rising contours.

Subsequently, we analyzed participants' RTs, with the same variables as in the Indonesian RT analyses. The results are summarized in Table 9. We found a main effect of *syllable*, which indicated that RTs were faster for words that contained the syllables [mae] and [lyy] compared to those that contained [dii]. Moreover, we found main effects of *type of contour* and of the quadratic and cubic terms of *step*. Inspection of the model's predicted scores showed faster RTs for falls than for rises. Further, participants responded relatively quickly near the left ends of the acoustic continua.

4.2. THE ZHUMADIAN TONE AND INTONATION DISCRIMINATION EXPERIMENT

The tone and intonation discrimination experiment was constructed by pairing stimuli from each of the (8 × 3 carrier syllables =) 24 continua. The members of the stimulus pairs were 1, 2, 3, and 4 steps apart, resulting in 18 stimulus pairs per continuum. We prepared two experiment files, A and B. Applying the selection criterion as for the Indonesian group, for file A we used the two sets of 18 pairs from the declarative tone continua for *dee*^{EF-LF} and

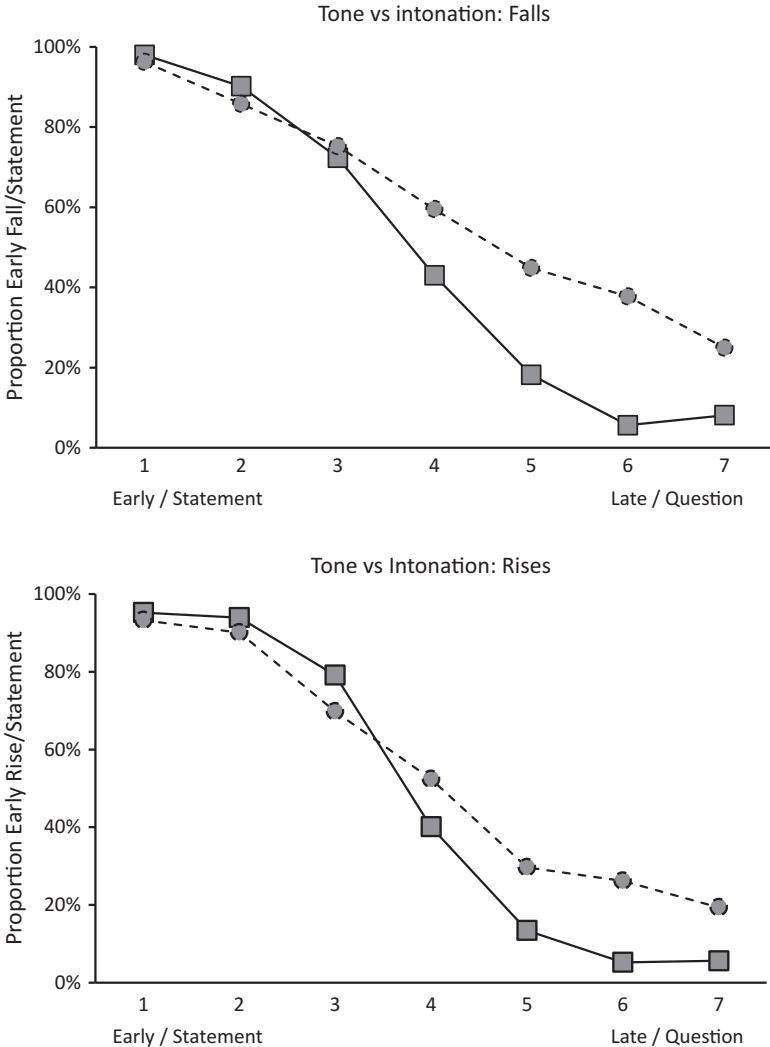


Fig. 6: Identification response curves for falls (top) and rises (bottom) by step for stimuli from intonation continua (---●---) and tone continua (—■—) by Zhumadian participants.

mae^{ER-LR} , and the two sets from the question tone continua for mae^{EF-LF} and lyy^{ER-LR} , a total of 72 stimulus pairs. For the intonational stimulus pairs, we selected the four sets of 18 pairs from the intonation continua for mae^{EF} , dee^{LF} , lyy^{ER} , and mae^{LR} , giving a total of 144 stimulus pairs for version A. Similarly, for B we selected the two sets of 18 pairs from the declarative tone continua for

TABLE 9. Results of the linear mixed-effects regression analysis of the logged RTs for the Zhumadian identification experiment

Random effects	Standard Deviation	
Listener	0.09	
Listener : Type of contour (rise)	0.04	
Fixed effects	Estimate	t
Intercept	1.441	98.24
Type of contour (rise)	0.041	5.47
Syllable (lyy)	-0.043	-5.49
Syllable (mae)	-0.026	-5.26
Step (linear)	-0.000	-0.03
Step (quadratic)	-0.015	-6.68
Step (cubic)	0.006	2.18

dee^{ER-LR} and *lyy*^{EF-LF} and the two sets from the question tone continua *mae*^{ER-LR} and *lyy*^{EF-LF}, while for the intonational stimulus pairs we selected the four sets from the intonation continua for *lyy*^{EF}, *lyy*^{LF}, *mae*^{ER}, and *dee*^{LR}, again giving 144 experimental pairs. To each of the two sets of 72 stimulus pairs, we added three filler pairs with the maximum step difference (i.e., steps 0 and 6), so that we had 150 pairs per experiment version. The order of the stimuli within pairs and the locations of the ‘same’ and ‘different’ boxes in version A, which was administered to 24 participants, were the reverse of those in version B, which was administered to 16 participants.

An instruction whereby ‘different’ could mean either ‘the same word, but a different intonation’ or ‘the same intonation, but a different word’, while ‘same’ would mean ‘same word and same intonation’ proved to be unworkable in a pilot presentation. We therefore presented the tone continua and the intonation continua in separate blocks. In the tone discrimination block, participants heard a pair of pronunciations which could be the same word or two different words and were asked to click either a box labelled ‘same word’ (相同的字) or a box labelled ‘different words’ (不同的字). In the intonation discrimination block, they heard a pair of pronunciations of the same word with either the same intonation or with different intonations and were asked to click either a box labelled ‘same intonation’ (相同语气的发音) or a box labelled ‘different intonation’ (不同语气的发音). The order of the stimuli within each stimulus pair was equally divided over the two directions of the step difference and the stimulus pairs in each block were randomly presented for each participant. A trial began with a 650 ms warning sound and a pause of 350 ms before

the first stimulus in the pair, followed after 400 ms by the second stimulus. In all, we collected 40×144 or 5,760 scores and reaction times, 2,880 for each of the two types of continua.

4.2.1. Results of the tone and intonation discrimination experiment: Zhumadian participants

Again expecting non-linear effects, we applied polynomial logistic mixed-effects regression analyses with the logit link function as for the Indonesian scores, including the same dependent and independent variables and a random intercept for *listener*. The results are shown in Table 10.

We found main effects of *syllable* and *step difference* as well as an interaction between *step difference* and *type of contrast*. Participants provided more ‘different’ responses for word pairs that contained the syllables [lyy] and [mae] compared to those that contained [dii], and, unsurprisingly, for word pairs with a larger step difference, which latter effect was larger in the case of the tonal stimulus pairs. The effects of *syllable* and *step difference* varied significantly across participants, as shown by the random slopes of these variables for *listener*.

Second, and more importantly, we found two-way interactions between *type of contrast* and the variables *step* and the quadratic term of *step*. To interpret

TABLE 10. Results of the logistic mixed-effects regression analysis for the Zhumadian tone and intonation discrimination experiment (0 = Same, 1 = Different)

Random effects	Standard Deviation		
Listener	0.83		
Listener : Step difference	0.29		
Listener : Syllable (lyy)	0.90		
Listener : Syllable (mae)	0.66		
Fixed effects	Estimate	z	p
Intercept	-1.439	-8.94	< .0001
Step difference	0.973	13.83	< .0001
Syllable (lyy)	0.425	2.55	< .05
Syllable (mae)	0.745	5.46	< .0001
Step (linear)	-0.223	-4.46	< .0001
Step (quadratic)	0.180	3.75	< .001
Type of contrast (tone)	1.095	11.72	< .0001
Step difference : type of contrast (tone)	0.259	3.48	< .001
Step (linear) : type of contrast (tone)	0.414	5.69	< .0001
Step (quadratic) : type of contrast (tone)	-0.583	-8.25	< .0001

these interactions, we split the data by *type of contrast*. For the tone discrimination scores, we found significant effects of *step* ($\beta = 0.212, z = 3.86, p < .001$) and *step (quadratic)* ($\beta = -0.434, z = -8.00, p < .0001$). The bottom panel in [Figure 7](#) shows that participants provided more ‘different’ responses for word pairs that were closer to the centre of the auditory continua, indicating a heightened sensitivity to phonetic differences between stimuli located immediately on either side of the tone continua midpoints. For the intonation discrimination scores, by contrast, we found significant effects of *step* and the quadratic term of *step* which were the reverse of those found for the tone discrimination data ($\beta = -0.232, z = -4.55, p < .0001$) and *step (quadratic)* ($\beta = 0.185, z = 3.79, p < .001$). The top panel of [Figure 7](#) shows that participants provided somewhat fewer ‘different’ responses for steps 4, 5, and 6 of the intonation continua, rather than somewhat more, as in the tone continua. Most crucially, the reverse effect for *step (quadratic)* indicates that intonation contrasts that were closer to the continuum midpoints were less perceivable to the Zhumadian participants than those closer to the endpoints, in particular those towards the Early/Statement ends (see [Figure 7](#)).

Subsequently, we analyzed participants’ RTs, measured from the appearance of the response buttons, as described in [Section 3.1.1](#). The results are summarized in [Table 11](#). The main effect of *syllable* indicates slower RTs for stimuli containing the syllable [mae] than for those with [dii], while the main effect for *type of contrast* indicates slower RTs for tone contrasts than intonation contrasts. Moreover, we found main effects of *step difference* and the quadratic term of *step difference*, indicating that RTs were faster for smaller step differences and this effect was strongest at the edges, i.e., for steps among stimuli closer to the forms of the language. The interaction between *step* and *type of contrast* is due to markedly faster RTs towards the left edge of the intonation continua than for their right edges and for tone continua generally.

4.3. DISCUSSION OF THE RESULTS OF THE ZHUMADIAN

IDENTIFICATION AND DISCRIMINATION EXPERIMENTS

Quite unlike the Indonesian participants, the Zhumadian participants produced sharply differing results for *type of contrast*. First, the identification results for the tone stimuli yielded a negative cubic effect of *step*, unlike the intonation contrast for the same group and unlike the tone and intonation results for the Indonesian group. That is, the Zhumadian scores on the tone continua uniquely yielded sigmoid curves for both rising and falling contours, indicating categoriality of the Early vs. Late tonal alignments. Also, the discrimination results for the tone contrasts uniquely yielded a large *negative* effect of the quadratic term for *step* in the Zhumadian group, unlike the intonation contrast for the same group and unlike the tone and intonation

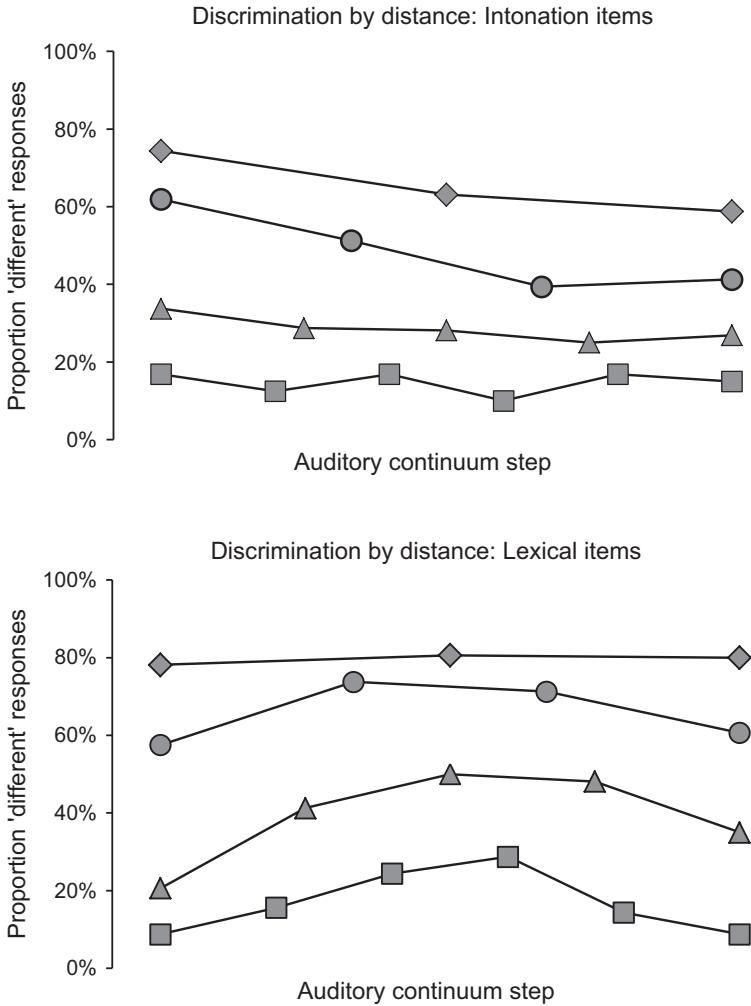


Fig. 7: Discrimination results for Zhumadian participants for intonation contrasts (top) and lexical contrasts (bottom) for 1 (■), 2 (▲), 3 (●), and 4 (◆)-step distances between members of pairs of stimuli.

contrasts for the Indonesian participants, all three of which yielded weaker *positive* effects. These results are comfortably in accordance with the predictions of our hypothesis.

A second difference between the two groups concerns the intra-group performance difference for *type of contrast*. For the Zhumadian participants,

CATEGORICAL TONE AND GRADIENT INTONATION

TABLE 11. *Results of the linear mixed-effects regression analysis of the logged RTs for the Zhumadian discrimination experiment*

Random effects	Standard Deviation	
Listener	0.07	
Listener : Step difference	0.01	
Listener : Syllable (l _{yy})	0.05	
Listener : Syllable (mae)	0.03	
Fixed effects	Estimate	<i>t</i>
Intercept	1.604	137.36
Step difference	0.008	2.80
Step difference (quadratic)	-0.005	-2.43
Syllable (l _{yy})	0.012	1.38
Syllable (mae)	0.036	4.90
Step (linear)	0.009	3.40
Step (quadratic)	0.003	1.31
Type of contrast (tone)	0.026	5.14
Step (linear) : type of contrast (tone)	-0.010	-2.73
Step (quadratic) : type of contrast (tone)	-0.005	-1.27

tones are more robustly identifiable and discriminable than intonations. This is shown by the larger number of misidentified carrier intonations than misidentified carrier words (see Figure 5). Additionally, the interaction between *step* and *type of contrast* in the Zhumadian identification scores shows how participants are less good at identifying question intonations, with scores at the question end not showing the same near-floor results as the near-ceiling results at the statement end (Figure 6, interrupted graphs in both panels), or than those for the lexical tone continua generally (Figure 6, uninterrupted graphs, both panels). This replicates earlier findings of the perceptual vulnerability of Mandarin question intonation in adults and children (Connell, Hogan, & Rozsypal, 1983; Singh & Chee, 2016; Xu & Mok, 2014), including the specific interaction of this vulnerability with lexical pitch movement direction, with rising tones increasing the vulnerability (Liu, Chen, & Schiller, 2016; Yuan, 2011). The overall better identifiability of stimuli from the tone continua does not appear to be attributable to a generalizable phonetic effect, because in the control group we found neither a main effect for *type of contrast* nor an interaction with *step*.

The RT effects for *syllable* in the experimental group are the reverse of those for the control group, with [dii] leading to slower response times for the Zhumadian participants but faster response times for the Indonesian participants, than [l_{yy}] and [mae]. This suggests that the continua for [l_{yy}] and [mae]

contained more language-specific phonetic properties than those for [dii] did, something which will have disadvantaged the Indonesian participants in their performance on [lyy] and [mae]. The lower number of ‘different’ responses by the Zhumadian participants for [dii] than for the other two syllables points in the same direction. The Indonesian participants’ relative advantage for [dii] is most probably enhanced by the presence of [di] and the absence of [y] and [ae] in their phonology (Section 3.3).

5. Conclusion and discussion

Despite their unusually subtle phonetic differences, the lexical tone contrasts in Zhumadian Mandarin between early and late aligning tones are categorical, i.e., discretely different, while the intonation contrast, both as implemented on falling lexical tones and as implemented on rising ones, is gradient. This finding is relevant to our understanding of human language as well as to the methodology we used to obtain it. As argued in the ‘Introduction’, this gradient-discrete distinction cannot motivate an exclusion of non-morphemic meanings from the scientific domain of language, but quite to the contrary implies that the two systems of communication must be studied together. We give three reasons for this.

First, while enjoying the benefits of the structural features of their languages, humans have evidently not abandoned the earlier direct signal-to-meaning approach to communication that exists in related species. An understanding of the structure of human language cannot be complete if we exclude these ‘pre-linguistic’ elements from our scientific perspective. For instance, we cannot currently say with any degree of confidence how common it is for languages to have intonational morphemes with statement and question meanings.

Second, non-arbitrary form–meaning correspondences frequently show up in uncontroversially morphemic shapes (Dingemanse, 2018). In our case, the non-arbitrariness is rooted in the speech production mechanism. For segmental affects, Ohala (1994) discussed the widespread occurrence of [i] and [j] elements in diminutive morphemes and claimed that their meanings derive from the short distance between the spread lips and the raised forward part of the tongue body, a gesture used to signal ‘smallness’ and related meanings like ‘submission’ and ‘friendliness’; Ohala (1980) proposed that this gesture is the origin of the smile, a non-aggressive signal specific to humans. He similarly associated question intonation with the projection of larynx size to body size, so that the higher pitches achieved in smaller larynxes, as in women, equally signal ‘smallness’ and related meanings like powerlessness (the Frequency Code). A variety of non-arbitrary intonational features based on other physiological factors have been claimed to develop into morphemes in many languages (Gussenhoven, 2016). Iconicity has not always lain behind the development

of non-structural signals into morphemes. Unsystematic cues of the hands, face, and head in the emerging Al-Sayyid Bedouin Sign Language grammaticalized into a prosodic system within three generations (Sandler, 2012; Sandler, Meir, Dachkovsky, Padden, & Aronoff, 2011). In a rigorous study of relative clauses in another young sign language, Israeli Sign Language, Svetlana Dachkovsky showed that manual and non-manual actions grammaticalized from unsystematic groupings and movements into consistent prosodic cues, in the second generation of signers of that language (Dachkovsky, 2017, to appear).

Such developments are part of our understanding of the ways in which the two systems of communication interact, but in the case of Mandarin it is conceivable that an earlier structural intonation system underwent a process of de-structuring. Our data are from a language with a high functional load for lexical tone which developed somewhere between the Old and Middle Chinese periods (Handel, 2015, p. 74). The competitive relation between lexical and intonational functions of pitch is supported by the impoverished intonation systems in varieties of Germanic with lexical tone, North Germanic Swedish/Norwegian and West Germanic Franconian (e.g., Gussenhoven & Peters, 2004; Kristoffersen, 2000, p. 100; Riad, 2014, p. 254), compared to the more complex intonation systems of the related non-tonal languages Icelandic and German (Árnason, 2011; Peters, 2018). The fact that Zhumadian Mandarin, like Standard Chinese, has a frequently used sentence-final question particle is therefore not coincidental, as indicated by the finding by Torreira, Roberts, and Hammarström (2014), who showed that languages with lexical tone have more syntactic devices for expressing interrogativity than languages without.

The third reason for a comprehensive study of morphemic and non-morphemic or not-quite-morphemic signals is that the meanings of the structural elements of intonational contours, i.e., the morphemes encoded as single or grouped intonational tones, have been framed as fairly skeletal operations by the speaker on the knowledge background presumed to be shared with the speaker (Brazil, 1975, 1985; Gunlogson, 2003; Gussenhoven, 2004 [1983]; Pierrehumbert & Hirschberg, 1990; Steedman, 2014). While much of the interpretation of the utterance will have to rely on inferences by the listener on the basis of his knowledge of the world, the sparse semantic information may well be crucially enhanced by the non-structural features in the intonation contour (Gili Fivela, 2012).

Importantly, we have shown that the classic categorical perception method, whereby responses from a discrimination task and an identification task with acoustic continua are jointly evaluated, can be used to address questions about the structural nature of a language's pitch contours. We avoided three problems that are associated with this type of experiment. One concerns the risk of comparing the results for continua with different degrees of acoustic gradualness, pointed out in the earliest literature on this topic (Section 1.3). All our

continua were acoustically comparable, since they ran between differently shaped falls or between differently shaped rises. The second problem is the absence of a cut-off point between results that allow for an interpretation of categoricity and those that indicate gradience (Gerrits & Schouten, 2004). We included baseline continua between uncontroversially discrete categories (tonal minimal pairs) by the side of continua that were hypothesized to be gradient (an intonation contrast). Third, it is conceivable that the results of any categorical perception experiment might not in fact be due to a language-specific effect, but to some quantal phonetic feature. We recruited a control group of participants, whose language uses none of the experimental phonetic differences in its phonology, whether lexical or intonational, so that their results could serve as a gradience baseline for all continua. This methodology goes beyond the collection of comparative data from native speakers of a non-tone language, whose responses may resemble those of the experimental tonal group, as shown by Gao, Tescano, Shih, and Tanner (2019), possibly because the lexical pitch under investigation appears as contrastive intonations in the non-tonal group. The robustness of our results is underscored by the fact that both the falls and the rises gave the same results. These two halves of our experiment can be regarded as an experiment and its replication for a functionally equivalent set of contours with a different overall pitch shape.

REFERENCES

- Abramson, A. S. & Lisker, L. (1970). Discriminability along the voicing continuum: cross language tests. *Proceedings of the sixth International Congress of Phonetic Sciences* (Vol. 196, No. 7, pp 569–573). Prague: Academia.
- Árnason, K. (2011). *The phonology of Icelandic and Faroese*. Oxford: Oxford University Press.
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bally, J. (1935). *Le langage et la vie*. Zürich: Max Niehans.
- Bao, Z. M. (1999). *The structure of tone*. New York/Oxford: Oxford University Press.
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1).
- Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.
- Best, C. T. (2019). The diversity of tone languages and the roles of pitch variation in non-tone languages: considerations for tone perception research. *Frontiers in Psychology* 10, e00364.
- Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer. Online <www.praat.org>.
- Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg (ed.). *Universals of human language: phonology: Vol. 2* (pp. 471–524). Stanford, CA: Stanford University Press.
- Borràs-Comes, J., Vanrell, M. & Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association* 44(1), 1–20.
- Brazil, D. (1975). *Discourse intonation* (Discourse Analysis Monographs 1). Birmingham: University of Birmingham, English Language Research.
- Brazil, D. (1985). *The communicative value of intonation in English*. Birmingham: Bleak House Books and English Language Research.
- Bruce, G. (1977). *Swedish word accent in sentence perspective*. Lund: Gleerup.

- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Chen, A., Gussenhoven, C. & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech* 47(4), 311–349.
- Chien, P.-J., Friederici, A. D., Hartswigsen, G. & Sammler, D. (2020). Neural correlates of intonation and tone in tonal and non-tonal language speakers. *Human Brain Mapping* 41(7), 1842–1858.
- Connell, B. A., Hogan, J. T. & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics* 11(4), 337–351.
- Cutting, J. E., Rosner, B. & Foard, C. (1976). Categories and boundaries in speech and music. *Quarterly Journal of Experimental Psychology* 28, 361–378.
- Dachkovsky, S. (2017). *Grammaticalization of intonation in Israeli sign language: from information structure to relative clause relations*. Unpublished PhD dissertation, University of Haifa.
- Dachkovsky, S. (to appear). From a demonstrative to a relative clause marker: grammaticalization of pointing signs in Israeli Sign Language. *Sign Language and Linguistics*.
- Dingemanse, M. (2018). Redrawing the margins of language: lessons from research on ideophones. *Glossa: A Journal of General Linguistics*, 3(1), 4, doi:10.5334/gjgl.444.
- DiCanio, C., Amith, J. D. & Castillo García, R. (2014). The phonetics of moraic alignment in Yoloxóchtitl Mixtec. In *TAL-2014*, 203–210. Online <<https://cupola.gettysburg.edu/anthfac/24/>>.
- Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E. & Peperkamp, S. (2008). Persistent stress ‘deafness’: the case of French learners of Spanish. *Cognition* 106(2), 682–706.
- Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language & Speech* 6(4), 206–217.
- Escudero, P., Benders, T. & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German, and Spanish listeners. *Journal of Phonetics* 37(4), 452–465.
- Fournier, R., Gussenhoven, C., Jensen, O. & Hagoort, P. (2010). Lateralization of tonal and intonational pitch processing: an MEG study. *Brain Research* 1328, 79–88.
- Fry, D. B., Abramson, D. B., Eimas, P. D. & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech* 5(4), 171–189.
- Gao, Y. A., Toscano, J. C., Shih, C. & Tanner, D. (2019). Reassessing the electrophysiological evidence for categorical perception of Mandarin lexical tone: ERP evidence from native and naïve non-native Mandarin listeners. *Attention, Perception, & Psychophysics* 81(2), 543–557.
- Gerrits, E. & Schouten, B. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics* 66(3), 363–376.
- Gili Fivela, B. (2013). Meanings, shades of meanings and prototypes of intonational categories. In G. Elordieta & P. Prieto (eds), *Prosody and meaning* (pp. 197–237). Berlin/New York: Mouton de Gruyter.
- Goedemans, R. & van Zanten, E. A. (2007). Stress and accent in Indonesian. In V. J. van Heuven & E. A. van Zanten (eds), *Prosody in Indonesian languages* (LOT Occasional Series 9) (pp. 35–62). Utrecht: Netherlands School of Linguistics.
- Grabe E., Gussenhoven, C., Haan, J., Marsi, E. & Post, B. (1998). Pre-accentual pitch and speaker attitudes in Dutch. *Language and Speech* 41(1), 63–85.
- Grice, M. & Baumann, S. (2007). An introduction to intonation-functions and models. In J. Trouvain & U. Gut (eds), *Non-native prosody: phonetic description and teaching practice* (pp. 25–51). Berlin/New York: De Gruyter.
- Gunlogson, C. (2003). *True to form: rising and falling declaratives as questions in English*. New York: Routledge.
- Gussenhoven, C. (2004 [1983]). *A semantic analysis of the nuclear tones of English*. Bloomington (Indiana): IULC.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C. (2006). Experimental approaches to establishing discreteness of intonational contrasts. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzyk, I. Mleinek, N.

- Richter & J. Schließer (eds), *Methods in empirical prosody research* (pp. 321–334). Berlin/New York: De Gruyter.
- Gussenhoven, C. (2016). Foundations of intonational meaning: anatomical and physiological factors. *Topics in Cognitive Science* 8(2), 425–434.
- Gussenhoven, C. & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43, 183–203.
- Gussenhoven, C. & Peters, J. (2004). A tonal analysis of Cologne Schärffung. *Phonology* 21(2), 251–285.
- Halim, A. (1974). *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Lembaga Bahasa Nasional.
- Hall, M.D. & Peck, R.B. (2017). Categorical perception: effects of the extent and rate of spectral change. *Journal of Cognitive Psychology* 29(1), 3–22.
- Handel, Z. (2015). Old Chinese phonology. In W. S.-Y. Wang & C. Sun (eds), *The Oxford handbook of Chinese linguistics* (pp. 65–69). Oxford: Oxford University Press.
- Hyman, L. M. (2011). Tone: Is it different? In J. A. Goldsmith, J. Riggle & A. C. L. Yu (eds), *The handbook of phonological theory* (2nd ed., pp. 179–239). Chichester: Wiley-Blackwell.
- Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H. & Tonoike, M. (1998). Task-dependent laterality for cue decoding during spoken language processing. *NeuroReport* 9, 899–903.
- Jaeger, F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4), 434–446.
- Kawahara, H., Masuda-Katsuse, I. & Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Communication* 27(3/4), 187–207.
- Kohler, K. (1987). Categorical pitch perception. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 331–333). Tallinn, Estonia. Online <https://www.ipds.uni-kiel.de/kjk/pub_exx/kk1987/tallinn-cat.pdf&usg=AOvVaw1uv5JFfyp2M0xWtIXvHrP1>.
- Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.
- Ladd, D. R. (2008 [1996]). *Intonational phonology*. Cambridge: Cambridge University Press.
- Liang, J. & van Heuven, V. J. (2007). Chinese tone and intonation in the perception by L1 and L2 speakers. In C. Gussenhoven & T. Riad (eds), *Tones and tunes. Volume 2: experimental studies in word and sentence prosody* (pp. 27–61). Berlin/New York: Mouton de Gruyter.
- Liberman A. M., Harris K. S., Hoffmann, H. S. & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54 (5), 358–368.
- Liberman A., Mattingly, I. G. & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (eds), *Coding processes and human memory*, (pp. 307–334). Washington, DC: Winston.
- Liu, F. & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62(2/4), 70–87.
- Liu, M., Chen, Y.Y. & Schiller, N. (2016). Online processing of tone and intonation in Mandarin: evidence from ERPs. *Neuropsychologia* 91, 307–317.
- Maskikit-Essed, R. & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology* 33(2), 353–389.
- Monroy, R. (2005). Spanish and English intonation patterns: a perceptual approach to attitudinal meaning. In C. S. Butler, M. de los Angeles Gómez González & S. M. Doval-Suárez (eds), *The dynamics of language use: functional and contrastive perspectives* (pp. 307–324). Amsterdam: Benjamins.
- Odé, C. (1994). On the perception of prominence in Indonesian. In C. Odé, V. J. van Heuven & E. van Zanten (eds), *Experimental studies of Indonesian prosody* (pp. 27–107). Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, University of Leiden.
- Odé, C. (2005). Neutralization or truncation? The perception of two Russian pitch accents on utterance-final syllables. *Speech Communication* 47(1/2), 71–79.

- Ohala, J. J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America* **68**, S33. doi:10.1121/1.2004679
- Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (eds), *Sound symbolism* (pp. 325–347). Cambridge: Cambridge University Press.
- Peperkamp S., Vendelin, I. & Dupoux, E. (2010). Perception of predictable stress: a cross-linguistic investigation. *Journal of Phonetics* **38**(3), 422–430.
- Peters, J. (2018). Phonological and semantic aspects of German intonation. *Linguistik Online* **88** (1). <https://doi.org/10.13092/lo.88.4191>
- Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. *Annual Review of Linguistics* **2**(1), 33–52.
- Pierrehumbert, J. B. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan & M. Pollack (eds), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Pierrehumbert, J. B. & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica* **46**(4), 181–196.
- Prieto, P. (2012). Experimental methods and paradigms for prosodic analysis. In A. C. Cohn, C. Fougeron & M. K. Huffman (eds), *The Oxford handbook of laboratory phonology* (pp. 528–537). Oxford: Oxford University Press.
- Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science* **6**(4), 371–381.
- Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
- Rahmani, H., Rietveld, T. & Gussenhoven, C. (2015). Stress “deafness” reveals absence of lexical marking of stress or tone in the adult grammar. *PLOS ONE*, **10**(12), e0143968.
- R Development Core Team, (2015). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Remijsen, B. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language* **89**, 297–327.
- Remijsen, B. & Ayoker, O. G. (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology* **31** (3), 435–462.
- Remijsen, B. & van Heuven, V. J. (1999). Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 2, pp. 1865–1868). Online <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1865.pdf>.
- Repp, B. H. (1984). Categorical perception: issues, methods, findings. In N. J. Lass (ed.), *Speech and language: advances in basic research and practice* (Vol. 10, pp. 244–322). Orlando, FL: Academic Press.
- Riad, T. (2014). *The phonology of Swedish*. Oxford: Oxford University Press.
- Roessig, S., Mücke, D., & Grice, M. (2019). The dynamics of intonation: categorical and continuous variation in an attractor-based model. *PLOS ONE*, **14**(5), e0216859.
- Sandler, W. (2012). Dedicated gestures in the emergence of sign language. *Gesture* **12**(3), 265–307.
- Sandler, W, Meir, I., Dachkovsky, S., Padden, C. & Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua* **121**(13), 2014–2033.
- Savino, M. & Grice, M. (2011). The perception of negative bias in Bari Italian questions. In S. Frota, P. Prieto & G. Elordieta (eds), *Prosodic categories: production, perception and comprehension* (pp. 187–206). London: Springer Verlag.
- Schneider, K. & Linfert, B. (2003). Categorical perception of boundary tones in German. In M. J. Solé, D. Recasens & J. Romero (eds), *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 631–634). Barcelona, Spain: Causal Productions Pty Ltd.
- Shen, G. & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: an ERP study. *Bilingualism: Language and Cognition* **22**, 253–265.

- Shen, X. S. (1990). *The prosody of Mandarin Chinese* (University of California Publications in Linguistics 118). Berkeley, BA: University of California Press.
- Singh, L. & Chee, M. (2016). Rise and fall: effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics* **55**, 109–118.
- Steedman, M. (2014). The surface compositional semantics of English intonation. *Language* **90**, 2–57.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics* **17**(1), 3–45.
- Stoel, R. (2006). The intonation of Banyumas Javanese. *Proceedings of Speech Prosody 2006* (pp. 827–830). Dresden: TUD press.
- Torreira, F., Roberts, S. G. & Hammarström, H. (2014). Functional trade-off between lexical tone and intonation: typological evidence from polar-question marking. In C. Gussenhoven, Y. Chen & D. Dediu (eds), *Proceedings of the 4th International Symposium on Tonal Aspects of Language* (pp. 100–103). <http://hdl.handle.net/11858/00-001M-0000-0018-1937-1>
- Vanrell, M. del M., Mascaró, I., Torres-Tamarit, F. & Prieto, P. (2012). Intonation as an encoder of speaker certainty: information and confirmation yes-no questions in Catalan. *Language and Speech* **56**(2), 163–190.
- Weitz, S. (1979). *Nonverbal communication: readings with commentary*. New York: Oxford University Press.
- Xi, J., Zhang, L., Shu, H., Zhang, Y. & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* **170**(1), 223–231.
- Xu, R. B. & Mok, P. P. K. (2014). Cross-linguistic perception of Mandarin intonation. In *Proceedings of the 7th International Conference on Speech Prosody* (pp. 638–642). Dublin. Online <https://www.isca-speech.org/archive/SpeechProsody_2014/pdfs/120.pdf>.
- Xu, Y. (2015). Intonation in Chinese. In W. S.-Y. Wang & C. Sun (eds), *The Oxford handbook of Chinese linguistics* (pp. 490–502). Oxford: Oxford University Press.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America* **130**(6), 4063–4069.
- Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X. & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLOS ONE* **6**(6), e20963.
- Zhang, Y. Y. (2018). *The two sides of Wh-indeterminates in Mandarin: a prosodic and processing account*. Utrecht: LOT.
- Zheng, H. Y., Minett, J. W., Peng, G. & Wang, W. S. (2012). The impact of tone systems on the categorical perception of lexical tones: an event-related potentials study. *Language and Cognitive Processes* **27**(2), 184–209.