Phonology **24** (2007) 187–214.
doi:10.1017/S0952675707001157 Printed in the United Kingdom

*Understanding near mergers: the case of morphological tone in Cantonese**

Alan C. L. Yu
University of Chicago

A growing body of work on exemplar-based theories of learning suggests the possibility of formal models of phonological representation which will offer deeper explanations of basic phonological properties than current models allow. The main purpose of this paper is to shed light on near merger, a recalcitrant problem in sound change and in phonological theory, with this newer perspective, through a case study of tonal near merger in Cantonese.

1 Introduction

A growing body of work on exemplar-based theories of learning suggests the possibility of formal models of phonological representation which will offer deeper explanations of basic phonological properties than current models allow. Exemplar-based models have proven useful not only in solving problems in phonetics and phonology (e.g. Pierrehumbert 1990, 2001, 2002, Goldinger 1996, Johnson 1997, Bybee 2001, Ernestus & Baayen 2003), but also in research domains as diverse as semantics, syntax, language acquisition and language change (see Gahl & Yu 2006 and references therein). The main purpose of this paper is to shed light from this newer perspective on near merger, a recalcitrant problem in sound change. Near merger describes the situation where speakers consistently report that two classes of sounds are 'the same', yet consistently differentiate them in production at better than chance level. Labov *et al.* (1972:

* During the preparation of this paper I benefited tremendously from discussions with (in alphabetical order) Rusty Barrett, Matt Goldrick, Bill Labov, Salikoko Mufwene and Janet Pierrehumbert. I would also like to thank audiences at the University of Pennsylvania, Northwestern University, University of Wisconsin, the Conference on Methods in Phonology at the University of California, Berkeley, and the Variation and Change in Phonology conference at Potsdam University. I would also like to thank Juliette Blevins and two anonymous reviewers for their discussion and criticism, which has led to considerable improvement of both the content and the presentation of my analysis. It goes without saying that any errors in this work are my own.

ch. 6), for example, find that speakers differentiate words like *source* and *sauce* in production, but report no distinction between them in perception. Similar near mergers have been reported in other varieties of English (e.g. *fool* and *full* in Albuquerque (Di Paolo 1988); *too vs. toe* and *beer vs. bear* in Norwich (Trudgill 1974); *line vs. loin* in Essex (Labov 1971, Nunberg 1980); *meat vs. mate* in Belfast (Milroy & Harris 1980, Harris 1985)). Many studies under the heading of 'incomplete neutralisation' have also reported small but consistent phonetic differences in supposedly neutralised environments. Thus, German has a fortis/lenis contrast between /t/ and /d/. This contrast occurs word-initially (e.g. *Tier* [tiːr] 'animal' *vs. dir* [diːr] 'to you') and intervocalically (e.g. *leiten* [laitən] 'lead' *vs. leiden* [laidən] 'suffer'), but only the fortis consonant occurs word-finally (see Iverson & Salmons 2007). Production and perceptual experimental results, however, show that the two sets of final voiceless stops are consistently different phonetically (Port & O'Dell 1985, Port & Crawford 1989). For example, the vowel or duration of closure voicing may be longer, or stop closure or burst shorter, in *Rad* /rad/ 'wheel' than in *Rat* /rat/ 'advice'. Similar findings have been reported for other languages, including Russian (Chen 1970, Burton & Robblee 1997), Catalan (Dinnsen & Charles-Luce 1984), Polish (Giannini & Cinque 1978, Tieszen 1997), Lezgian (Yu 2004) and Dutch (Warner *et al.* 2004).

The main puzzle of near merger and incomplete neutralisation is how speakers manage to maintain a systematic production difference when they consistently fail to identify the distinction at the conscious or near-conscious level. This is especially perplexing within traditional theories of the phonetics–phonology interface (Cohn 1990, Keating 1990, Pierrehumbert 1990), which assume that phonological representations in the lexicon are categorical, contrastive elements, while the phonetic implementation component computes the degree and timing of articulatory gestures, which are gradient and variable. Information flow is strictly unidirectional in the sense that no articulatory plan can look backward to the phonological encoding, nor can the phonological encoding look back to the lexical level. No lexical information can influence the phonetic implementation directly either, bypassing the level of phonological encoding. On this view, the categorical form of a lexeme wholly determines the phonetic outcome. Phonetic variations on the surface are considered artefacts of the context or performance-induced anomalies. The discovery of *systematic* subphonemic differences between representations that are otherwise taken to be identical is unexpected. As Labov *et al.* (1991: 38) summarise:

> The assumption that contrasts were discrete and binary, that there was no such thing as a small difference in sound, that production and perception were symmetrical, and that introspections were reliable, all militated against such an idea [i.e. near merger – AY].

From this perspective, near merger and incomplete neutralisation, which could collectively be referred to as 'suspended contrasts', are essentially two sides of the same coin. That is, both phenomena involve an unexpected distinction at the phonetic level between lexical items (in the case of near merger) or phonemes (in the case of incomplete neutralisation), which escapes detection by traditional methodologies of phonological investigation. The main distinction between near merger and incomplete neutralisation has to do with the purported site of contrast elimination. Near merger is found at the level of the lexicon or the morphology and is therefore context-free. Incomplete neutralisation is context-dependent, since the location of contrast neutralisation is restricted to a phonologically defined context.

Reports of incomplete neutralisation and near mergers have puzzled linguists of all stripes. Responses to the existence of incomplete neutralisation and near merger vary widely in the literature. Some phonologists categorically deny the existence of the phenomenon (e.g. Manaster Ramer 1996). Others suggest that the subtle phonetic differences observed are better explained as a consequence of orthographic differences or as variation in speaking style. For example, it has been found that the less the experimental design emphasises the role of orthography, the smaller the durational effects (Fourakis & Iverson 1984, Jassem & Richter 1989). Port & Crawford (1989) find that discriminant analysis to classify productions by underlying final voicing is most successful (78 % correct) when speakers dictate the words, but least successful (approximately 55 % correct) when target words are embedded in sentences that do not draw attention to the minimal pairs (whether read or repeated orally). Focusing on the case of final devoicing in Dutch, Warner *et al.* (2004) control for possible orthographic influence and still find subphonemic durational differences in that language. They also show that listeners can perceive durational differences which are not consistently observed in production. Ernestus & Baayen (2003) find evidence that listeners can use these subphonemic distinctions to hypothesise which past tense allomorph non-words would take.

While the phonological community reacted to the existence of 'suspended contrasts' with caution, the reactions from the sociolinguistic community proved to be much more receptive. Labov (1975), for example, concludes that the existence of near merger suggests that sound change may bring two phonemes into such close approximation that semantic contrasts between them are suspended for native speakers of the dialect, without necessarily leading to merger. Largely unresolved is the question of what mechanism, grammatical or otherwise, leads to this type of 'suspension of contrasts'. In near merger, the underlying category difference has to be supported by something, e.g. contact with another dialect that maintains the distinction (Labov 1994) or, in the case of literate cultures, by orthographic differences (Faber & Di Paolo 1995).

The goal of this paper is twofold. The main empirical contribution comes from a case study of tonal near merger in Cantonese (§2). The morphologically derived mid-rising tone in Cantonese shows a fundamental frequency profile consistently distinct from that of the lexical mid-rising tone. Taking the Cantonese tonal near merger as a jumping-off point, this paper offers a theory of near merger couched within an exemplar-based model of speech production and perception. Near merger is explained as the by-product of extensive overlapping of the exemplar clouds of two or more categories. This exemplar-based view of near merger finds further support from the origins of tonal near merger in Cantonese (§4). Discussion and concluding remarks appear in §5.

2 Tonal morphology in Cantonese

Cantonese has six tone types. Traditional Chinese philology treats syllables with final stops (checked syllables) as distinct tone classes (checked tones), yielding a nine-tone system. Until recently, there was also a contrast between high level and high falling; however, this distinction has collapsed for most speakers today. Open syllables are generally treated as phonemically CV, since there is no phonemic vowel-length contrast in Cantonese, even though phonetically they are more accurately represented as CVV.

(1)

Unchecked tones			*Checked tones*		
55 (~53)	si	‘poetry’	55	sɪk˺	‘to know’
33	si	‘to try’	33	sɛk˺	‘to kiss’
22	si	‘affairs’	22	sɪk˺	‘to eat’
23	si	‘market’			
35	si	‘to cause, make’			
21	si	‘time’			

Chen (2000)’s analysis of the tonal system of Cantonese, which assumes that there are three tone heights, H, M and L, is adopted here. Contour tones are represented as sequences of level tones. A summary of the phonological representation of the Cantonese tone system, adapted from Chen (2000: 33), is given in Table I. For ease of reference, the corresponding Chao tone number is given in superscript. Chen uses ‘q’ to indicate the checked tones.

	level			
	CV(N)	CVq	rising	falling
high (*yin*) low (*yang*)	M^{33} L^{22}	H^5q, M^3q L^2q	MH^{35} LM^{23}	HM^{53} (~H^{55}) ML^{21}

Table I
The phonological representation of the Cantonese tone system (adapted from Chen 2000: 33).

This section begins with a presentation of a set of phenomena in the tonal morphology of Cantonese. A production experiment is reported on in §2.2, and a perception experiment in §2.3.

2.1 The 'changed tone' in Cantonese

Other than the lexical mid-rising tone (henceforth R^{lex}), mid-rising tones may also come about as the result of a set of morphological alternations, which are collectively referred to as *pinjam*, or 'changed tone', in the Cantonese linguistic literature. Following Downer (1959) and Kam (1977), *pinjam* may be regarded as manifestations of a system where certain morphemes are derived from their respective semantic correlates just by a change of tone. There are two phonetically distinct products of this tone change. Historically high-falling tone (i.e. 53) changes to high-level (i.e. 55). However, since the high-falling tone is for the most part no longer contrastive, this alternation is irrelevant to the phonologies of most speakers. The only clear case of *pinjam* is the derivation of mid-rising tones from semantically related syllables with a non-high-level, non-mid-rising tone. For example, the nominalisation of verbs carrying a non-rising tone (2a) is indicated purely by a change in tone (2b).

(2) a. *Level tone* b. *Rising tone*

sou33	'to sweep'	→	sou35	'a broom'
pɔŋ22	'to weigh'	→	pɔŋ35	'a scale'
mɔ11	'to grind'	→	mɔ35	'a grind'
tɑn22	'to pluck'	→	tɑn35	'a missile'
wɑ22	'to listen'	→	wɑ35	'an utterance'
jɐu11	'to grease'	→	jɐu35	'oil'
liu11	'to provoke'	→	liu35	'a stir'
tsʰɵɥ11	'to hammer'	→	tsʰɵɥ35	'a hammer'
tsʰɔ11	'to plough'	→	tsʰɔ35	'a plough'

The productivity of the *pinjam* pattern (henceforth R^{morph} for 'morphologically derived rising tone') is unclear.[1] Tonal alternations are largely unpredictable. Thus, the /jɐn11/ 'human' element in /nɵɥ23 jɐn35/ 'female' and /nɑn11 jɐn35/ 'male' has a *pinjam*, while the same element in /kuŋ55 jɐn11/ 'labourer' and /nuŋ11 jɐn11/ 'farmer' does not. While the condition for the application of *pinjam* appears random, nonetheless, several generalisations hold. Several morphological contexts give rise to *pinjam* (e.g. vocative, adjective reduplication, etc.). The alternant generally denotes the nominalisation of a verbal action (2) or a familiar and/or diminutive object (3a) and various other complicated semantic nuances (Jurafsky 1988). Wong (1982) reports that words with *pinjam* like those in (3b) occur more frequently in less formal situations than in

[1] As discussed in §4, mid-rising tones may also come about as a result of a set of sandhi patterns.

higher-register speech. Kam (1977) reports that many native speakers do not recognise any relationship between the derived forms and their alleged base. It should be noted that the number of these *pinjam* derived syllables is sizeable. For a comprehensive discussion of the phonology and semantics of *pinjam*, see Bauer & Benedict (1997: ch. 2).

(3) a.	t^{h}oi21	'stage, terrace'	→	t^{h}oi35	'table'
	k^{wh}ɐn21	'skirt'	→	wɐi21 k^{wh}ɐn35	'apron'
	kɛŋ33	'mirror'	→	ŋɑn23 kɛŋ35	'eyeglasses'
	nɵɥ23	'girl, woman'	→	moʊ23 nɵɥ35	'call girl'
	nɵɥ23	'girl, woman'	→	sɐʊ55 nɵɥ35	'nun'

b.	kɑt̚22 tsɑt̚22	~	kɑt̚22 tsɑt̚35	'cockroach'
	wu21 tip̚22	~	wu21 tip̚35	'butterfly'
	ɑp̚33	~	ɑp̚35	'duck'
	kɑp̚33	~	kɑp̚35	'pigeon, squab'

Previous phonological treatments advocate a floating high tone morpheme account of *pinjam* (Yip 1980, 2002, Chen 2000). The floating high tone attaches to the end of the relevant syllable and creates a new 35 tone (4).

```
(4) sou  'to sweep'        sou  'a broom'
     |                  →   |‾‾‾‾\
     M  <H>                 M     H
```

While there are several morphological patterns that give rise to R^{morph}, since they are all analysed phonologically as the docking of a floating high tone onto the host syllable, the term *pinjam* or R^{morph} will continue to be used as a cover term to describe these morphological alternations as a whole.

Given the fact that the alternation in tone signifies a change in meaning, *pinjam* is necessarily a morpholexical pattern (i.e. not a postlexical phenomenon). Thus floating tone association must take place within the lexicon, as is generally assumed under a modular feed-forward model of phonetics and phonology interface. Neutralisation within the lexical phonology component is absolute (i.e. Structure Preservation; Kiparsky 1985), thus the prediction of such a model is that segmentally identical syllables with R^{lex} and R^{morph} respectively are implemented in the same way by the phonetic implementation component, since the underlying morphological difference between these two tones does not persist past the phonological component to the phonetic implementation component. Let us refer to this as the Absolute Neutralisation Hypothesis.

(5) *Absolute Neutralisation Hypothesis*

Identical phonological surface representations have similar phonetic realisations.

Experiment 1, which investigates the phonetic realisation of *pinjam* and R^{lex}, was conducted to ascertain the validity of this prediction.

Cantonese tone has been the focus of many phonetic studies (e.g. Fok 1974, Vance 1976, 1977, Vance & Walker 1976, Gandour 1981, Kong 1987, Bauer & Benedict 1997, Wong & Diehl 1998, 1999, Wong 1999, Flynn 2003, Francis & Ciocca 2003, Francis *et al.* 2003). Some studies have also discussed the phonetic aspects of *pinjam* in Cantonese. Cheng (1973) finds that *pinjam* in the Taishan dialect of Cantonese has a higher F0 maximum than the highest level tone in the language. Taishan Cantonese has five basic tones, three level tones and two falling tones (6). Cheng uses a seven-point scale to draw attention to the exact spacing of the tones and the extra-high F0 maximum of *pinjam*.

(6)

Basic tone	*Pinjam*
66	*no change*
44	447
22	227
52	527
31	317

Cheng's detailed study of the pitch tracking shows that *pinjam* in Taishan Cantonese has the shape of the basic tones, with an added rise to the very high tone at the end. However, the significance of the extra height of *pinjam* is unclear, since Taishan Cantonese has no lexical rising tone. Thus, unlike Standard Cantonese, Taishan *pinjam* is structure-building in the sense that it introduces additional tones to the language. No merger is involved. With respect to Standard Cantonese, Whitaker (1955–56: 188) and Yuan (1983) notice impressionistically that the contour of R^{morph} rises higher than the contour of R^{lex}. Based on the production of six speakers from Guangzhou, Hong Kong and Macau, Bauer & Benedict (1997) report no significant differences between the realisation of the R^{lex} and R^{morph} tones. However, it should be noted that, while the actual F0 values were reported for each speaker in their study, no statistical analysis was attempted. Taking these reports into account, Flynn (2003: 64) concludes that *pinjam* should not be treated as an independent tone, since no systematic experimental evidence has ever been provided in support of the claim that *pinjam* in Standard Cantonese has a different tonal value. Experiment 1 is designed to ascertain the phonetic correlates of *pinjam*, and how *pinjam* compares to lexical mid-rising tone.

2.2 Experiment 1: a production study

2.2.1 *Methods.* **Subjects:** Six native speakers of Hong Kong Cantonese (three males and three females), all undergraduates at the University of Chicago, were paid a nominal fee for participating in the production experiment. They were all born and raised in Hong Kong. With the exception of one of the male speakers (EK), who had been in the United States for eight years, the other subjects had been in the United States for less than two years. They spoke English as a second language. None of them reported any speech or hearing problems.

Stimuli and procedures: The target CVN syllables, underlined in (7), are ten (near) minimal pairs (i.e. with identical rhyme). Each target syllable has two versions (R^{lex} *vs.* R^{morph}).

(7)

Lexical 35		*Derived* 35	
søŋ55 fɑn35	'opposite'	kɑm55 fɑn35	'prisoner'
kɐi55 tɑn35	'egg'	fɛi55 tɑn35	'a missile'
pow55 pin35	'to critique'	tsʰɵɥ21 pin35	'casual'
suŋ55 pɔŋ35	'to untie'	tsʰuŋ23 pɔŋ35	'heavy'
tsʰøŋ21 kɛŋ35	'long neck'	ŋɑn23 kɛŋ35	'glasses'
fɑ55 fɐn35	'pollen'	ku35 fɐn35	'(stock) share'
tsʰɵt̚55 pɑn35	'to publish'	siu35 fɑn35	'peddler'
tsiː22 tin35	'dictionary'	tʰiu21 kin35	'terms, conditions'

To minimise the subjects' awareness of the focus of the experiment, the target disyllabic words/phrases of this experiment are embedded and randomised within a larger set of tokens of different syllable structure and different tone types. The subjects recited the list of Cantonese disyllabic words/phrases written in the carrier phrase /ŋɔ23 wui23 tuk̚22 __ pɛi35 nɛi23 tʰæŋ55/ 'I am going to say __ for you to hear' three times. Each subject recited, in total, 300 sentences (i.e. 10 target tokens × 2 versions × 3 repetitions + 80 fillers × 3 repetitions = 300 tokens). The randomised tokens were divided into three lists. Subjects were given the opportunity to take breaks between lists. Subjects' productions were recorded in a sound-proofed room, using a Marantz PMD670 solid state recorder and a SONY C-48 condenser microphone, which was placed approximately six inches from the subject's mouth.

Acoustic measurement: The speech was digitised at a sampling rate of 48,000 Hz, and all measurements were made with Praat (Boersma & Weenink 2005). The following parameters were measured: the F0 at the onset of the sonorous phase of the syllable, the F0 peak of the rising contour, the duration of the sonorous phase of the syllable and the duration of the F0 peak relative to the onset of voicing. Recordings of the Cantonese contour tone syllables showed a slightly falling F0, occurring over the first quarter to the first third of the syllable (cf. Bauer & Benedict 1997, Vance 1977, Flynn 2003). The point at which the slope changes is referred to here as the inflection point (IP; Francis *et al.* 2003). Since the inflection point might interact with the realisation of the F0 peak, it was estimated visually from the F0 plot, and both the time (relative to the onset of voicing) and F0 at that point were also measured. Only tokens with clear F0 contours (i.e. tokens not affected by excess creakiness) were included in the results reported below.

2.2.2 *Results*. Outliers, defined as F0 more than two standard deviations from the means of their condition, were replaced by condition means.

This occurred for 1·5% of the onset measurements, 1·8% of the inflection-point measurements and 2% of the peak measurements. A within-subject design was used to examine the surface tone realisation with different underlying tonal specification: R^{lex} and R^{morph}. Data were analysed using a one-way repeated measures ANOVA. There was a significant main effect of underlying tone type at all three points of measurements (Onset: $F(1,5)=9{\cdot}42$, $p=0{\cdot}028$; IP: $F(1,5)=21{\cdot}62$, $p=0{\cdot}006$; Peak: $F(1,5)=28{\cdot}37$, $p=0{\cdot}003$). Thus the underlying tonal specification affects the surface realisation of the rising tone.

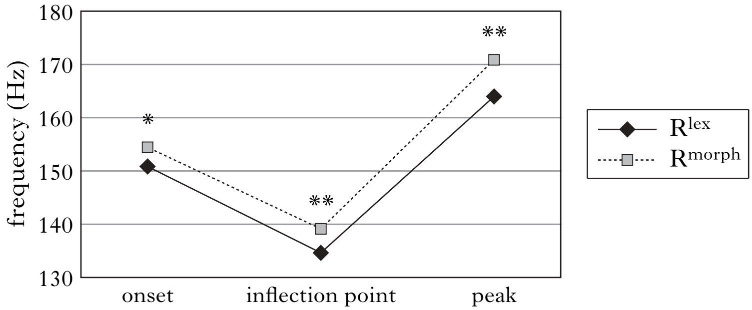
	speaker	R^{lex}	R^{morph}	df =	F	p
onset	AK	106·46	113·27	(1, 59)	15·96	0·000**
	CS	187·72	189·88	(1, 56)	0·75	0·390
	EK	121·02	124·28	(1, 59)	1·83	0·182
	JK	172·37	179·82	(1, 46)	3·79	0·058
	MC	206·90	208·36	(1, 59)	0·15	0·703
	VC	111·55	111·03	(1, 47)	0·05	0·824
inflection point	AK	95·35	100·79	(1, 59)	11·37	0·001**
	CS	167·73	169·23	(1, 55)	1·03	0·314
	EK	108·43	114·24	(1, 58)	17·66	0·000**
	JK	153·34	160·64	(1, 47)	10·02	0·003**
	MC	179·23	184·47	(1, 58)	7·00	0·011*
	VC	103·72	105·46	(1, 47)	1·10	0·300
peak	AK	115·55	117·55	(1, 59)	8·65	0·005**
	CS	207·11	219·81	(1, 57)	13·59	0·001**
	EK	156·33	164·26	(1, 59)	7·44	0·008**
	JK	176·55	182·67	(1, 47)	5·69	0·021*
	MC	210·35	216·44	(1, 59)	5·18	0·027*
	VC	118·09	121·88	(1, 47)	3·96	0·053

Table II

Mean F0 values of R^{lex} and R^{morph} at three temporal locations (the onset of voicing, the inflection point and the F0 peak), and the ANOVA results for the F0 difference between R^{lex} and R^{morph}.

* significant at $p<0{\cdot}05$, ** significant at $p<0{\cdot}01$.

Tables II and III show the results of the acoustic measurements by individual subjects. Table II summarises the F0 measurements, showing that, with the exception of speaker VC, the F0 maxima of the two rising tones are significantly different from each other, contrary to the prediction of the Absolute Neutralisation Hypothesis (5). In particular, the F0 maximum of R^{morph} is higher than that of R^{lex}. Four of the six speakers show a significant difference in the F0 of the inflection point as well.



	onset	inflection point	peak
R^{lex}	150·84	134·63	163·99
R^{morph}	154·44	139·13	170·87

F0 position

Figure 1

Mean F0 values of R^{lex} and R^{morph} at three temporal locations (the onset of voicing, the inflection point and the F0 peak). * significant at $p<0{\cdot}05$, ** significant at $p<0{\cdot}01$.

As shown in Fig. 1, the overall contour of R^{morph} is higher than the contour of R^{lex}. In particular, the F0 values of the inflection point and the peak of the rise are generally higher for R^{morph} than R^{lex}. The results of a one-way ANOVA for the rate of change of the rising contour beginning at the inflection point for the two rising tone types show no significant difference.[2] This suggests that the overall rise is the same between the two rising tone types, but the overall contour is shifted upwards in the case of R^{morph}.

Table III summarises the durational measurements. No significant durational difference is found between the R^{lex} and R^{morph} syllables, except for speaker MC, who shows a difference in the overall duration of the sonorous phase of the syllable and the time it takes for the initial portion of the F0 contour to reach the inflection point.

2.2.3 *Discussion.* The results of Experiment 1 indicate that the F0 contours of R^{lex} and R^{morph} are significantly different, confirming Whitaker and Yuan's impressionistic observations. One interpretation of the production results is that there are two sets of mid-rising tone in Standard Cantonese, akin to the analysis of Taishan *pinjam* mentioned earlier. The insensitive phonetic transcription by previous researchers has therefore obscured a tonal distinction in the language. Previous perceptual studies have found that Cantonese speakers can reliably identify and discriminate contrastive tones in the language (Fok 1974, Vance 1977, Francis *et al.* 2003). Thus, if R^{morph} were indeed a contrastive tonal category in Cantonese, one would also expect that native speakers of Cantonese would

[2] The rate of change is calculated using the following formula: $(Dur_{F0peak} - Dur_{IP})/F0_{Peak} - F0_{IP})$.

	$Duration_T$			$Duration_{IP}$			$Duration_{PEAK}$		
speaker	diff	df	F	diff	df	F	diff	df	F
AK	−9·67	(1,59)	2·45	−1·97	(1,59)	0·36	−3·90	(1,59)	0·67
CS	0·79	(1,58)	0·01	−27·54	(1,55)	0·89	5·48	(1,58)	0·47
EK	0·16	(1,59)	0·00	−9·37	(1,59)	3·01	1·12	(1,59)	0·02
JK	−10·12	(1,47)	2·08	−1·76	(1,47)	0·12	−5·79	(1,47)	0·71
MC	−17·52*	(1,59)	4·32	−10·84*	(1,58)	4·19	−12·76	(1,59)	2·30
VC	−12·31	(1,47)	2·93	−4·30	(1,47)	0·75	0·04	(1,47)	0·00

Table III

Mean differences and results of the ANOVAs of three durational measures: the overall duration of the sonorous phase of the syllable ($Duration_T$), the time it takes for the F0 fall to reach the inflection point ($Duration_{IP}$), the time it takes for the F0 rise to reach its peak ($Duration_{PEAK}$).

* significant at $p<0{\cdot}05$.

be capable of reliably distinguishing between R^{lex} and R^{morph}. To test this hypothesis, a small-scale identification experiment was conducted.

2.3 Experiment 2: an identification experiment

2.3.1 *Methods.* **Subjects:** Five native speakers of Hong Kong Cantonese (two males and three females), all undergraduates at the University of Chicago, were paid a nominal fee for participating in this identification experiment. Two of the participants (CS and AK) also participated in the first experiment.

Stimuli and procedures: The tokens of speaker EK for the production experiment were chosen for the perceptual test, since his tokens showed the greatest difference between R^{lex} and R^{morph}. The stimuli were a set of five minimal pairs, shown in (8), which consisted of target syllables with R^{lex} or R^{morph}.

(8) *Lexical* 35 / *Derived* 35

fɑn35	'opposite'	fɑn35	'prisoner'
fɐn35	'pollen'	fɐn35	'(stock) share'
kɛŋ35	'long neck'	kɛŋ35	'glasses'
pin35	'to critique'	pin35	'casual'
pɔŋ35	'to untie'	pɔŋ35	'heavy'

30 tokens (10 syllables × 3 repetitions) were used as stimuli in the identification test. An experimental design program, E-Prime, was used to administer the experiment. The tokens were placed in random order by E-Prime, and the subjects were given a ten-alternative forced-choice task. The stimuli were played to the listeners through headphones as they sat alone in a quiet room. After a stimulus was played, the listeners were

		responses									
		fan$_{M}$	fan$_{L}$	fɐn$_{M}$	fɐn$_{L}$	kɛŋ$_{M}$	kɛŋ$_{L}$	piːn$_{M}$	piːn$_{L}$	pɔŋ$_{M}$	pɔŋ$_{L}$
stimuli	fan$_{M}$	8	7								
	fan$_{L}$	10	4		1						
	fɐn$_{M}$			3	12						
	fɐn$_{L}$			4	11						
	kɛŋ$_{M}$					3	12				
	kɛŋ$_{L}$					3	12				
	piːn$_{M}$							8	7		
	piːn$_{L}$						1	6	8		
	pɔŋ$_{M}$						1			3	11
	pɔŋ$_{L}$									5	10

Table IV

The number and distribution of identification responses by five listeners. Subscript M = R^{morph}; subscript L = R^{lex}. The stimuli are arranged by rows, while the responses are given in columns.

presented with ten response alternatives written in Chinese characters on a computer screen, and identified the word by clicking on the appropriate Chinese character.

2.3.2 *Results.* The confusion matrix of the responses is given in Table IV. The stimuli are arranged in rows, while the responses are arranged in columns. The mean correct identification rate is 46·7 %. While the average correct identification rate is above chance level (10 %), the listeners' identification of R^{lex} and R^{morph} remains quite poor. The rate of the listeners misidentifying one type of rising tone for the other is 51·3 %, suggesting that the listeners are confusing R^{lex} and R^{morph}. It should, however, be noted that the most common responses seemed to be words with lexical tone, since approximately 63 % of the responses were for words with lexical tones, while only 35 % of the responses were for derived words.

2.3.3 *Discussion.* The results of the production and identification experiments establish that the phonetics of morphological tone in Cantonese reflects a near-merger scenario. That is, while speakers show distinct tonal realisations for R^{lex} and R^{morph} in production, listeners are not able to identify the appropriate lexical item based on the different tonal information alone. The simple introduction of an additional tonal category in Cantonese is therefore unsatisfactory. Further qualification is needed to explain the apparent near merger between R^{lex} and R^{morph}. The fact that R^{morph} occurs only in a limited, morphologically definable set of environments demands an explanation as well.

The discovery of near merger in Cantonese is significant for several reasons. To begin with, phonetic evidence for near merger involving tonal features has never been reported before. Thus, the addition of tonal contrasts to the set of 'suspendable phonetic contrasts' strengthens the empirical support for near merger as a genuine phenomenon of language. Cantonese tonal near merger is similar to other reported cases of near merger in which listeners consistently produce a small difference between two contrasting words without being able to reliably recognise the difference in either their own or other people's productions. However, there are notable differences between Cantonese tonal near merger and previously reported cases of suspended contrasts. In traditional cases of incomplete neutralisation such as German syllable-final devoicing, the underlying category difference is supported by an alternation at the phonemic level. That is, the underlying category difference can be recovered paradigmatically, since the underlying voicing category is recoverable in certain phonologically definable environments. The underlying category difference between R^{lex} and R^{morph}, on the other hand, is supported by alternations at the morphological level. That is, the underlying category difference cannot be recovered from the immediate phonetic environment where the respective tone type occurs.

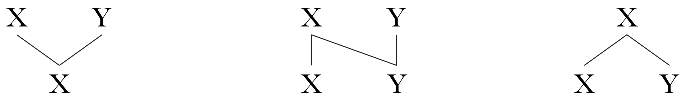
It is also important to point out that these results cannot simply be dismissed as an artefact of the language studied, since Cantonese speakers do not possess conscious knowledge of the tonal phonology or the morphology of the language. That is, Cantonese has no standard orthography or accepted Romanisation that is taught in school. To the extent that a modified version of Standard Chinese orthography is used by Cantonese speakers, that orthography still does not indicate tonal information, nor does it reflect morphological changes. Thus the source of the observed near merger must rest elsewhere; orthography *per se* offers no support for the suspension of complete merger in this case. In §4, several theories of the origins of tonal near merger are discussed. As will be demonstrated, tonal near merger is the result of variations introduced by the morphology of the language. Cantonese tonal near merger thus provides an instructive example of near merger that is not sustained by dialect contact or orthographic factors. But before discussing of the origins of *pinjam*, the next section will articulate a theory of near merger.

3 Sound change as exemplar redistribution

As noted in the introduction, traditional modular feed-forward models of phonetics and phonology predict that the categorical form of the lexeme wholly determines the phonetic outcome. According to the modular feed-forward model, two words should be implemented similarly if they share the same phonological representation. The model of the lexicon and the phonetics–phonology interface advocated in this paper diverges from such models in several important ways. An exemplar-based speech-processing system recognises inputs and generates outputs by analogical evaluation across a lexicon of distinct memory traces of each token of speech. A category label is defined by a 'cloud' of exemplars (i.e. memorised tokens). Each exemplar may contribute to many categories simultaneously. While specific exemplar models differ in how new experiences are assigned to relevant categories and how that new experience is integrated with the stored exemplars, all exemplar-based models assume that each experience alters the entire category system slightly by changing the range, and perhaps activation, of component exemplars (Hintzman 1986, Nosofsky 1986, Kruschke 1992, Goldinger 1996). As the perceptual memories associated with the label (i.e. the exemplar cloud) accumulate and are incrementally updated, the distribution of these forms may shift. To account for recency effects, most exemplar-based models assume some form of time-sensitive base-activation decay (Pierrehumbert 2001). Furthermore, exemplar models can also accommodate abstract categorical behaviours via abstract generalisations that can be continually created and updated through experience (Tenpenny 1995). This model has received an impressive array of empirical support in recent years. Studies in perception have shown that listeners may rely on fine-grained phonetic details in word recognition (e.g. Goldinger 1996, Johnson 1997, Ernestus & Baayen 2003). Pierrehumbert (2002) reviews a large body of literature that shows that allophonic details are systematically associated with words. Clopper & Pisoni (2004) show that dialect perception appears to be exemplar-based as well. These cases are difficult to explain without assuming that individual words have associated phonetic distributions.

Recent years have also witnessed much growth in applying exemplar-based models to the understanding of language change. For example, Bybee (2001) reviews evidence from leniting historical changes, arguing that lexical representations of words must include incrementally updated information about the phonetic distribution of each word. Kirchner (1999) offers a computational simulation of phonologisation using an exemplar-based speech processing system. Batali (2002) investigates the emergence of recursive syntactic structures using an exemplar-based computational simulation of language acquisition. Focusing on the problem of sound change, exemplar-based models conceive of changes in the sound system as the result of shifts and redistribution of exemplars within the exemplar space. Within this model of sound change, changes in the pronunciation norm are largely involuntary and fully dependent on the listener's linguistic experience. Following Pierrehumbert (2001), the production of a category label results from a random selection of a category exemplar correlated with the probability of the base-activation level of an exemplar. The production target is the result of averaging some number of exemplars within a region of an exemplar cloud (Pierrehumbert 2002). This averaging between exemplars predicts a steady convergence to the mean over the entire category. When the exemplar cloud of one category is mixed with the exemplars of another, the averaging procedure might recruit exemplars of both categories. Over time, the means of the two categories will gradually converge. Three major types of phonological change thus find natural expressions within this model, as shown in (9).

(9) a. *unconditional merger* b. *conditioned merger/ primary split* c. *secondary split*



An unconditional merger (9a) obtains when an exemplar cloud shifts its distribution to overlap completely with another. A secondary split (9c) takes place when a unimodal distribution of exemplars becomes bimodal or multimodal. Conditioned merger and primary split (9b) obtain when a unimodal distribution of exemplars becomes bimodal and one of the resultant exemplar clouds overlaps completely with another exemplar cloud. In these cases, mergers are complete, because exemplars previously classified under one category end up being classified under a different category.

On this view, near merger represents the transition stage in a complete merger scenario. That is, near merger results when the exemplar clouds of two or more categories overlap sufficiently but the category membership of the exemplars remain distinct. Since exemplars are categorised in any number of salient dimensions, under most communicative situations a given phonetically ambiguous token may still be categorised accurately in other dimensions, based on contextual and indexical information. Listeners have troubles identifying the appropriate category label in a laboratory or fieldwork setting, since important contextual cues are often absent (see also Labov 1994: ch. 14). When forced to make a categorical judgement based solely on the acoustic information, identification is at chance level, since the perceived utterance falls within the combined exemplar cloud. The production of a near-merged category, however, is no different from any other. It proceeds by creating a target based on the average of some number of exemplars, including those within the region of the partially overlapping exemplar clouds. The mean production difference between two merging categories might be small, but they will continue to show distinct phonetic realisation as long as the category means remain separate.

Returning to the case of Cantonese tonal near merger, we hypothesise that the morphological tone category is the result of incomplete mergers between two or more tones in the immediate past. The main question that must be addressed, however, is what tones were merged and how the production mean of the merged tones gave rise to the slightly higher fundamental frequency profile of the Changed Tone. The answer lies in the origins of the Changed Tone, which is the focus of the next section.

4 Explaining Cantonese tonal near merger

It is widely recognised that Cantonese *pinjam* comes from the elision of certain morphemes that are no longer synchronically productive. Several possible sources have been proposed in the literature. Beginning with Chao (1948), many have noted that *pinjam* is probably the cognate of *er*-suffixation in other Chinese dialects (Whitaker 1955–56, Wong 1982). While the modern pronunciation of Mandarin *er* 'child' in Cantonese is /ji21/ 'child', according to Wong (1982), an alternative pronunciation of the word, /ji55/, must have been prevalent at the time when *pinjam* developed. The two pronunciations possibly represent a difference in speech style. This situation is not uncommon, as the Cantonese lexicon is divided into roughly two layers: the literary and the colloquial registers. Reflexes of this now lost colloquial morpheme /ji55/ can be found in certain lexicalised forms in Modern Cantonese (10).

(10)	mɑu55	'cat'	*vs.*	mɑu55 ji55	'kitten'
	hɑt̚55	'to beg'	*vs.*	hɑt̚55 ji55	'beggar'

Based on dialectal evidence, Bauer & Benedict (1997) propose that Cantonese *pinjam* must have developed from a monosyllable that carried a rising tone. For example, in the Bobai dialect of Yue spoken in the southeastern Guangxi Province, the more conservative variety employed the diminutive suffix [ɲin] with a long rising tone: /kae32 ɲin25/ 'small chicken', /iaŋ21 ɲin25/ 'small sheep', /ma21 ɲin25/ 'small horse' (Wang 1932, Wong 1982, Bauer & Benedict 1997). When words ended in a final stop, the final stop would change to its homorganic nasal counterpart and the word would acquire a long rising tone, associated with /ɲin25/ (11). According to Bauer & Benedict, what is found in Modern Cantonese today reflects the stage where the final nasal suffix has been dropped and *pinjam* remains as the sole marker of the morphological change.

(11)	œk54	'house'	→	oŋ25	'little house'
	mat32	'thing'	→	man25	'little thing'
	hɔp4	'box'	→	ham25	'little box'
	pak1	'uncle'	→	a33-paŋ25	'my little uncle'

The morpheme /tsi35/ 'son' has also been advanced as a possible source of *pinjam* (Cheng 1973, Wong 1982). Unlike the morpheme /ji55/ in (10), the use of /tsi35/ to mark diminutive is still semi-productive today. Many of the /tsi35/-attached forms also have a *pinjam* alternative (12).

(12)	min22 tsi35	~	min35	'face, honour'
	kʰɛi21 tsi35	~	kʰɛi35	'a chess piece'
	jɐt̚2 tsi35	~	jɐt̚35	'(special) day'

To summarise, *pinjam* originated from the loss of certain diminutive morphemes. Several candidates have been proposed: /ji55/ 'child',

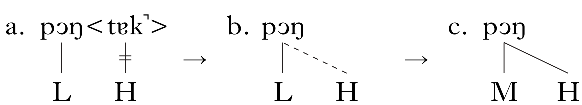
/ɲin25/ 'child' and /tsi35/ 'son'. While proposals as to the specific origins of *pinjam* vary, crucially, these morphemes are all realised with a high or rising tone. The generalisation that emerges from the historical evidence is clear: the tonal morpheme today is a relic of an earlier syllable-contraction phenomenon. When the morpheme in question elides, the orphaned high or rising tone is realised on the preceding syllable. *Pinjam* thus represents the morphologisation of a sandhi pattern in the immediate past. From the point of view of the present theory of sound change, the historical scenario suggests that the fundamental frequency profile of *pinjam* is the reflex of combined fundamental frequency profiles of a host of sandhi tones. One question remained unanswered by the above historical scenario, however. Why is *pinjam* realised differently from the lexical rising tone? While the precise phonetics of the sandhi pattern that gave rise to *pinjam* is lost for ever, a set of sandhi patterns still active in Cantonese today provides us with important clues on how the sandhi tone is realised at the pre-morphologised *pinjam* stage.

4.1 Experiment 3: a phonetic study of contracted syllables

Several modal elements in Cantonese which have either an underlying high or mid-rising tone often elide in colloquial speech, causing the tone of the preceding syllable to change to mid-rising. Two types of elision or contraction are illustrated in (13). (13a) shows that the perfective marker, /tsɔ/, merges with the head verb. (13b) shows the merger of the potential marker, /tɐk˺/, with the head verb. The perfective marker carries a lexical 35 tone, while the potential marker has a lexical 55 tone. The two types of resulting merged syllables are traditionally transcribed identically with a 35 tone.

(13)	a.	pɑŋ22 tsɔ35	→	pɔŋ35	'to weigh (PERF)'
		fɑn22 tsɔ35	→	fɑn35	'to transgress (PERF)'
		kin33 tsɔ35	→	kin35	'to meet (PERF)'
	b.	pɔŋ22 tɐk˺55	→	pɔŋ35	'to weigh (POTENTIAL)'
		fɑn22 tɐk˺55	→	fɑn35	'to transgress (POTENTIAL)'
		kin33 tɐk˺55	→	kin35	'to meet (POTENTIAL)'

Phonologically, this syllable-contraction pattern (also referred to as a sandhi phenomenon in the literature) is often characterised as the re-association with the preceding syllable of a high tone (14b), set afloat by the elision of an originally high tone or rising tone syllable (Yip 1980, Chen 2000). A rule of register readjustment gives rise to the appropriate representation of the MH cluster (14c).

(14)


Thus, if *pinjam* were indeed the reflex of a historical syllable-contraction phenomenon similar to those found in Modern Cantonese (13), the phonetic realisation of the modern-day sandhi rising tone should shed light on the source of the *pinjam* F0 profile as well. To this end, an experiment was conducted to examine the acoustics of the sandhi tones in (13). While the results are quite variable, they nonetheless offer some clues as to the likely origins of the subphonemic difference observed in Experiment 1.

4.1.1 *Methods.* **Subjects:** The same six speakers who participated in Experiment 1 also participated in this experiment.

Stimuli and procedures: Ten pairs of verb + modal phrases in Cantonese were used. The list of verbs used, all CVN in shape, is given in (15). Two modal markers were used: /tɐk̚55/ (POTENTIAL) and /tsɔ35/ (PERFECTIVE). After the application of syllable elision, the two phrases in each pair should be homophonous. For ease of reference, the rising tone resulting from the elided /tɐk̚55/ will be referred to as RS1 from now on, while the rising tone resulted from the elided /tsɔ35/ will be referred to as RS2.

(15)	hɑŋ21	'to walk'	tɐn22	'to warm up, simmer'
	tɑn22	'to bounce, pluck'	pɔŋ22	'to weigh'
	fɑn22	'to offend'	pʰiːn33	'to lie'
	tiːn22	'to electrify'	kiːn33	'to see'
	pɑn22	'to pretend'	tɛŋ33	'to throw'

Subjects were instructed to recite a randomised set of the target phrases written in the carrier phrase /ŋɔ23 wui23 tuk̚22 __ pɛi35 nɛi23 tʰæŋ55/ three times in 'lazy speech style', to maximise the likelihood of syllable elision. Each subject recited a total of 60 sentences (i.e. 10 target tokens × 3 repetitions × 2 lists = 60 tokens). Subjects were given the opportunity to take breaks between lists. The recording procedure was the same as for Experiment 1.

4.1.2 *Acoustic measurement.* The results of the experiment are quite mixed. Only three speakers (EK, JK, VC) produced some form of syllable contraction. The other three speakers either refused to acknowledge the possibility of contraction in Cantonese at all (Speakers AK, CS) or produced no noticeable contraction in the recitation (Speaker MC). Of the three speakers who did produce syllable contraction 'on the fly', only JK and VC show actual syllable elision. While Speaker EK produced syllable contraction, the modals were not entirely elided. This result is akin to the type of syllable-fusion phenomenon reported in Wong (1996, 2004), where the second syllable is obscured by onset lenition or deletion and vowel reduction – the target phrase thus remains disyllabic (e.g. /pɔŋ22 tsɔ35/ → /pɔ22 ŋɔ35/). Since the focus here is on the realisation of RS, non fully elided examples of syllable contraction will not be considered further. Only tokens with complete second (modal) syllable elision

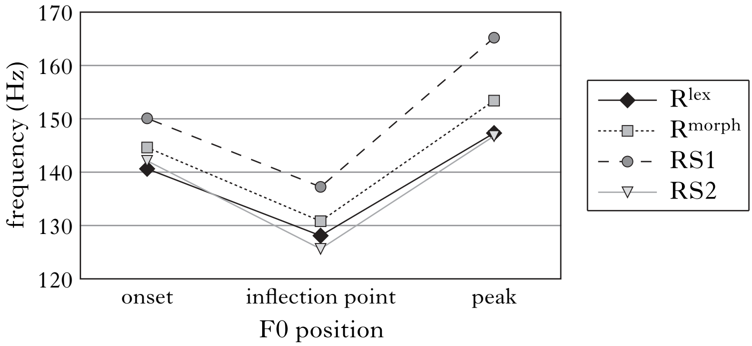


Figure 2
Mean F0 values of R^{lex}, R^{morph}, RS1 and RS2 by the two speaker at three temporal locations (the onset of voicing, the inflection point and the F0 peak).

(e.g. /pɔŋ22 tsɔ35/ → /pɔŋ35/) or those fused with the targeted syllable as a short vocalic release of the final nasal (e.g. /pɔŋ22 tsɔ35/ → /pɔŋᵊ35/) were taken into account.

The same acoustic measurements performed in Experiment 1 were taken, again using Praat. Only tokens with clear F0 contours (i.e. tokens not affected by excess creakiness) were included in the results reported.

4.1.3 *Results*. Outliers, defined as F0 values more than two standard deviations from the means of their condition, were again replaced by condition means. This occurred for 1% for the inflection point measurements and 4% of the peak measurements. Since both JK and VC participated in the first experiment, their R^{lex} and R^{morph} production results were used for comparisons with the RS1 and RS2 data. A summary of the measurements is given in Fig. 2. Data were analysed using a one-way ANOVA with four levels (R^{lex}, R^{morph}, RS1 and RS2). There were significant main effects of the derived status of a tone at the F0 onset ($F(3, 189) = 2{\cdot}933$, $p < 0{\cdot}05$), inflection point ($F(3, 190) = 3{\cdot}675$, $p < 0{\cdot}05$) and F0 peak ($F(3, 190) = 7{\cdot}808$, $p < 0{\cdot}01$). Thus the derived status of a tone affects the surface realisation of the rising tone.

The results of a series of post hoc pairwise comparison are given in Table V. The tonal contour of R^{lex} is not significantly different from that of R^{morph} or that of RS2.[3] The F0 peak of RS1 is significantly higher than that of R^{lex} and R^{morph}. RS1 and RS2 are significantly different at all three points of measurement. The overall duration of the sonorous phase of the

[3] The lack of a significant difference between R^{lex} and R^{morph} is consistent with the results of Experiment 1. That is, Experiment 1 reveals that all subjects except VC show significant differences in the production of R^{lex} and R^{morph}. Since the R^{lex} and R^{morph} data considered in Experiment 3 are pooled from only two subjects, JC and VC, it is not surprising that the differences between R^{lex} and R^{morph} do not reach significance.

	onset difference	IP difference	peak difference
$R^{lex} - R^{morph}$	−4·63	−4·97	−4·96
R^{lex} – RS1	−9·48	−8·64	−16·66*
R^{lex} – RS2	11·87	10·77	13·73
R^{morph} – RS1	−4·85	−3·67	−11·70*
R^{morph} – RS2	16·50	15·73*	18·69*
RS1 – RS2	21·35*	19·40*	30·39**

Table V

F0 differences based on pairwise comparison of R^{lex}, R^{morph}, RS1 and RS2 at three temporal locations (the onset of voicing, the inflection point and the F0 peak). * significant at p<0·05, ** significant at p<0·01.

RS syllables is generally significantly longer than those of the R^{morph} and R^{lex} syllables. The RS1 syllables (VC = 237·3 ms; JC = 233·51 ms) are on average about 15–20 % longer than the R^{morph} and R^{lex} syllables, while the RS2 syllables (VC = 267·02 ms; JC = 254·49 ms) are on average about 35–40 % longer than the R^{morph} and R^{lex} syllables.

4.2 Discussion

The F0 contours of the RS tones are not completely neutralised with the R^{morph} F0 contour. The direction of incomplete neutralisation, however, differs, depending on the nature of the fused syllable. The tonal contour of RS1 is found to be higher than R^{morph}, but the tonal contour of RS2 is lower. On the other hand, RS2 and R^{lex} share similar F0 values, while the F0 peak of RS1 is significantly different from R^{lex}. The F0 profiles of RS1 and RS2 do not match as well; RS1 has a significantly higher rising profile than that of RS2. The incomplete merger between RS1 and RS2 seems puzzling at first glance. Why should the two types of fused syllable behave differently in terms of their F0 contours? Here, the difference between the F0 contours of RSI and RS2 can be explained as a difference between the elided morphemes: the perfective marker /tsɔ/ carries an underlying 35 tone, while the potential marker /tɐk̚/ has an underlying 55 tone. The phonological analyses in (14) and (16) help clarify this point. Syllable contraction and the resulting tonal fusion are treated as part of the postlexical phonology of Cantonese. The rising contour of the perfective marker, /tsɔ/, is analysed as a sequence of level tones under a single tonal root note, abbreviated in (16) as R (Yip 1989).[4] RS2 is the result of R reassociating with the preceding host syllable (16b). A set of tonal

[4] This single tonal root analysis is in contrast with what Yip refers to as a tonal cluster, which is a sequence of two tonal root nodes associated with a single syllable.

readjustment rules prevent surface sequences of a level tone and a rising tone, so that the underlying L tone of the verb drops out in (16c).

(16) a. pɔŋ<tsɔ> → b. pɔŋ → c. pɔŋ
L R   L R   R

Phonetically speaking, the fact that RS2 has a similar tonal profile to R^{lex} is not surprising, since the perfective marker /tsɔ/ is expected to have a lexical mid-rising tone profile. All else being equal, the fused syllable should retain the F0 contour of the elided syllable. A similar explanation can be advanced for RS1. Phonologically, the RS1 syllable results from the disassociation of the high tone from the potential marker /tɐk˺/, as in (14a) above. As shown there, the unassociated H then docks onto the preceding host syllable, and a register-readjustment rule gives rise to the MH cluster in (14c).

As Lee (1999) reports, the F0 of a 55 tone in a checked syllable is higher than the peak of the F0 rise of a lexical 35 tone for some speakers. Ladefoged (2000) also transcribes the mid-rising tone as 24 and the high tone as 55, suggesting that the peak of the rising tone is slightly lower than that of a high tone. Thus, the extra-high F0 of the potential marker can be interpreted as the retention of the tonal profile of an underlying 55 tone.

The phonetic realisation of the RS tones revealed in Experiment 3 offers instructive clues as to why *pinjam* today has a tonal profile different from that of R^{lex}. Recall that *pinjam* was the result of the elision of a set of morphemes with inherent high or mid-rising tone, similar to the RS tones examined above. Assuming the uniformitarian hypothesis, we may hypothesise that the phonetic realisation of the sandhi tones prior to the emergence of the *pinjam* category shared similar phonetic qualities with the RS1 and RS2 tones today. As shown in Experiment 3, the tonal contour of a sandhi tone varies, depending on the tone of elided syllable. That is, an elided high tone syllable gives rise to a higher rising contour than an elided rising tone syllable does. In an exemplar-based model of speech perception and production, the production of a phonological representation of a category involves making a selection from the exemplar cloud for that category. As the perceptual memories associated with the category accumulate and are incrementally updated, the distribution of these forms may shift. Thus, when the sandhi patterns lose their productivity and become morphologised, the exemplars of what would have corresponded to modern-day RS1 and RS2 tones would come to represent a unified category. The mean of this combined category would therefore lie somewhere between the extremes of the two sandhi tone exemplar clouds. Figure 3, which shows the distribution of the F0 peak values of RS1 and RS2 on the one hand, and the distribution of the F0 peak values of R^{morph} on the other, sketches how this would give rise to a distinct tonal profile for the *pinjam* category. Based on the data obtained in Experiment 3, we know that the mean probability distributions of the F0 peak values of RS1

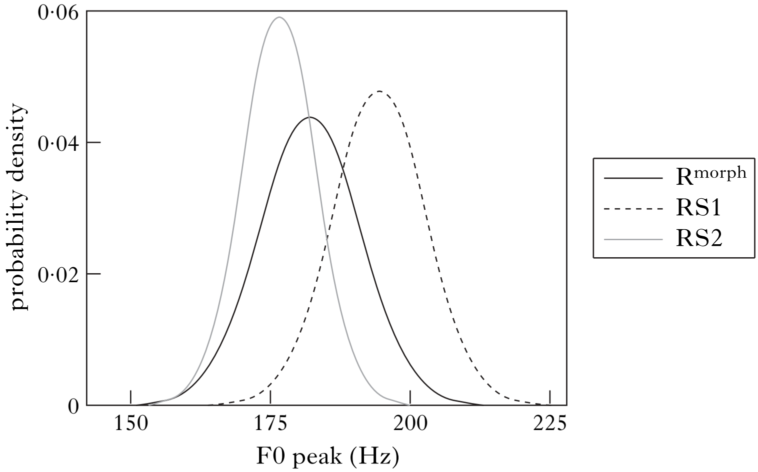


Figure 3

Total frequency distributions for F0 peak values of R^{morph}, RS1 and RS2.

and RS2 fall at opposite ends of the F0 continuum. If these two exemplar clouds were combined to form a new category, as arguably happened in the origins of the *pinjam* category, the production of this combined category would draw its exemplars from a diffused exemplar cloud. Given the averaging procedure mentioned above, the production target of the combined category would fall somewhere between the two extremes of the F0 continuum. As the resulting production tokens were stored as new exemplars in their category of origin, the perception-production feedback loop would result in exemplars steadily populating the exemplar region corresponding to the mean over the entire category (cf. Wedel 2007). As illustrated in Fig. 3, the distribution of R^{morph} lies between the RS1 and RS2 distribution, and is thus consistent with the predictions of the present model. Note that, on this exemplar-based understanding of *pinjam*, there is no need to posit a new distinct tonal contrast in Cantonese. The fact that the R^{morph} exemplars pattern together emerges from their morphological patterning (i.e. only exemplars of the types deverbal noun, diminutive and vocative may carry *pinjam*). The class behaviour of R^{morph} cannot be established on the basis of phonological distribution alone. This exemplar-based understanding thus obviates the need to fit *pinjam* into some parochial representation such as a single branching tone unit or a tonal cluster. To the extent that representational analyses of the sort in (14) and (16) are useful for conceptualising abstract relations between exemplars, *pinjam* may be understood as the conglomeration of exemplars of both tonal clusters and branching tone units (i.e. what would correspond to modern-day RS1 and RS2 respectively).

The present model also predicts incomplete neutralisation between R^{morph} and R^{lex}. Recall that the tonal profile of RS2, which resulted from elided mid-rising tone syllables, does not significantly differ from that of R^{lex}. Thus, given that the mean distribution of R^{morph} lies above the mean of RS2, the mean distribution of R^{morph} should lie above R^{lex} as well. Listeners cannot reliably identify R^{morph} from R^{lex}, because their distributions overlap extensively with one another.

The current state of R^{morph} can thus be seen as the result of a two-stage development. In the first stage, a set of syllable-contraction patterns gave rise to two types of phonetically distinct sandhi tones, one with a higher F0 peak than the other. As these sandhi alternations became morphologised, a new morphological category, i.e. the *pinjam* tone, was established. The R^{morph} exemplar cloud thus encompasses the exemplar spaces of both types of sandhi tones. A new production norm for *pinjam* gradually emerged as new *pinjam* exemplars steadily populated the region of exemplar space which corresponds to the mean of the entire *pinjam* exemplar cloud. The R^{morph} and R^{lex} exemplar clouds do not merge completely, since the mean of the R^{morph} exemplar cloud is skewed by the many high F0 exemplars, which are relics of sandhi tones that resulted from fused high tone morphemes. R^{lex} would lack such exemplars intrinsically.

5 Conclusion

The existence of near merger has puzzled phonologists and historical linguists for many years. Using a case study of tonal near merger as a jumping-off point, this paper draws on insights from a growing body of work on exemplar-based models of phonological representation and offers a concrete model of sound change. This model relies on the notion that phonological representations are viewed as aggregates of exemplars of linguistic experiences, rather than a set of abstract categorical units. Changes in the sound system are the result of shifts in the organisation of the exemplar space. When two or more exemplar clouds of different categories overlap extensively, listeners are predicted to have difficulty in reliably identifying tokens falling within the combined exemplar space. The production difference between near-merged tokens is predicted to be small, due to the proximity of exemplars being drawn for calculating the production target. Previous literature suggests that the underlying category difference in a near-merger situation may be supported by contact with another dialect that maintains the distinction (Labov 1994) or, in the case of literate cultures, by orthographic differences (Faber & Di Paolo 1995). This paper argues that underlying category difference in a near-merger situation can be sustained by grammar-internal factors as well. That is, the production difference between lexical and morphological rising tones is sustained by the morphology/semantics alone.

The model of sound change articulated in §3 not only provides a principled explanation for near merger, but also offers useful insights for problems unresolved in previous models of sound change. For example, many have assumed that the typological convergence of sound change results from language-universal factors, such as physiological and psychological factors common to all human speakers at any time period (Ohala 1993, Blevins 2004). Sound change takes place when the listener mistakes the effects of the speaker's production system, of ambient effects on the acoustic stream or of his or her own perceptual system as representative of the speaker's internal representations or computations.

Within this comprehensive model of sound change, taking into account production, perception and feedback effects of both on the developing grammar, a potential source of sound change are 'errors' that listeners make in (re)constructing pronunciation norms (Ohala 1993) based on pronunciation variation (cf. 'CHOICE' in Blevins 2004). While much effort has been devoted to understanding the nature of misapprehension, what is less clear is why listeners would ever change their assumption about the language in the face of seemingly haphazard evidence of deviation from the norm. At issue here, therefore, is the nature of the pronunciation norm (re)construction. One way to account for the eventual change in pronunciation norms is to assume that the listener may not have the experience to enable him to do the necessary normalisation or correction. Two special groups lacking this experience are children acquiring phonological systems for the first time and adult second-language learners (Ohala 1993: 247). Within this model, primary agents of change are individuals with a near blank slate, due to their lack of previous significant exposure to the language being learned.

The assumption that children and second language learners are the only agents of sound change is, however, too restrictive, particularly in light of a growing body of research demonstrating that it is possible for adults to participate in sound change, albeit less readily than children (Sankoff 2004). For example, Harrington *et al.* (2000) compare the vowel space of Queen Elizabeth II, based on her annual Christmas broadcasts spanning three time periods (the 1950s, the late 1960s/early 1970s and 1980s), and find that her vowel system has gradually shifted toward a more mainstream form of Received Pronunciation. Sancier & Fowler (1997) investigate the voice onset time (VOT) production of a native Brazilian Portuguese speaker, and find that the speaker's VOTs for Brazilian Portuguese and English stops are shorter after a stay of several months in Brazil than after a similar stay in the United States. The findings of the Harrington *et al.* and Sancier & Fowler studies demonstrate that a speaker well past the critical period for language acquisition may change his pronunciation norm upon prolonged exposure to the ambient language. Learning, as it were, is ongoing. In light of the evidence, how can the apparent need to include adults in the set of potential participants in linguistic change be reconciled with the listener-misperception view of sound change? The answer lies in the proper understanding of what constitutes the PRONUNCIATION NORMS. Within the model of sound change argued for in this paper, changes in the pronunciation norm are largely involuntary and dependent on the listener's linguistic experience.

The 'blank slate' problem vanishes in this model, since no assumption about the linguistic experience of the agent of change is explicitly required. The fact that adults may participate in sound change is explained by the fact that they share with children the same mechanism of speech production and perception. However, other factors, such as the critical period of acquisition or the impoverished linguistic experience of a child (i.e. fewer competing exemplars) may explain why sound change can occur, overtly or covertly, when children learn to speak.

REFERENCES

Batali, John (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Ted Briscoe (ed.) *Linguistic evolution through language acquisition*. Cambridge: Cambridge University Press. 111–172.

Bauer, Robert S. & Paul K. Benedict (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Blevins, Juliette (2004). *Evolutionary Phonology: the emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul & David Weenink (2005). *Praat: a system for doing phonetics by computer (version 4.3.14)*. http://www.praat.org.

Burton, Martha W. & Karen E. Robblee (1997). A phonetic analysis of voicing assimilation in Russian. *JPh* **25**. 97–114.

Bybee, Joan (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Chao, Yuen Ren (1948). *Mandarin primer*. Cambridge, Mass.: Harvard University Press.

Chen, Matthew Y. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* **22**. 129–159.

Chen, Matthew Y. (2000). *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Cheng, Teresa M. (1973). The phonology of Taishan. *Journal of Chinese Linguistics* **1**. 256–322.

Clopper, Cynthia & David B. Pisoni (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *JPh* **32**. 111–140.

Cohn, Abigail (1990). *Phonetic and phonological rules of nasalization*. PhD dissertation, University of California, Los Angeles.

Di Paolo, Marianna (1988). Pronunciation and categorization in sound change. In K. Ferrara, B. Brown, K. Walters & J. Baugh (eds.) *Linguistic change and contact: Proceedings of the 16th Annual Conference on New Ways of Analyzing Variation in Language*. Austin: University of Texas. 84–92.

Dinnsen, Daniel A. & Jan Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. *JPh* **12**. 49–60.

Downer, G. B. (1959). Derivation by tone-change in Classical Chinese. *Bulletin of the School of Oriental and African Studies* **22**. 35–78.

Ernestus, Mirjam & R. Harald Baayen (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Lg* **79**. 5–38.

Faber, Alice & Marianna Di Paolo (1995). The discriminability of nearly merged sounds. *Language Variation and Change* **7**. 35–78.

Flynn, Choi-Yeung-Chang (2003). *Intonation in Cantonese*. Munich: Lincom Europa.

Fok, Chan Yuen-yuen (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong.

Fourakis, Marios & Gregory K. Iverson (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica* **41**. 140–149.

Francis, Alexander & Valter Ciocca (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *JASA* **114**. 1611–1621.
Francis, Alexander L., Valter Ciocca & Brenda Kei Chit Ng (2003). On the (non) categorical perception of lexical tones. *Perception and Psychophysics* **65**. 1029–1044.
Gahl, Susanne & Alan C. L. Yu (eds.) (2006). Special theme issue: Exemplar-based models in linguistics. *The Linguistic Review* **23**. 213–379.
Gandour, Jack (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics* **9**. 20–36.
Giannini, Antonella & Umberto Cinque (1978). *Phonetic status and phonemic function of the final devoiced stops in Polish*. Naples: Seminario di studi dell'Europa orientale.
Goldinger, Stephan D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**. 1166–1183.
Harrington, Jonathan, Sallyanne Palethorpe & Catherine Watson (2000). Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association* **30**. 63–78.
Harris, John (1985). *Phonological variation and change: studies in Hiberno-English*. Cambridge: Cambridge University Press.
Hintzman, Douglas L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review* **93**. 411–428.
Iverson, Gregory K. & Joseph C. Salmons (2007). Domains and directionality in the evolution of German final fortition. *Phonology* **24**. 121–145 (this issue).
Jassem, Wiktor & Lutosława Richter (1989). Neutralization of voicing in Polish obstruents. *JPh* **17**. 317–325.
Johnson, Keith (1997). Speech perception without speaker normalization: an exemplar model. In Keith Johnson & John W. Mullennix (eds.) *Talker variability in speech processing*. San Diego: Academic Press. 145–165.
Jurafsky, Daniel (1988). On the semantics of Cantonese changed tone: women, matches, and Chinese broccoli. *BLS* **14**. 304–318.
Kam, Tak Him (1977). Derivation by tone change in Cantonese: a preliminary survey. *Journal of Chinese Linguistics* **5**. 186–210.
Keating, Patricia A. (1990). Phonetic representation in a generative grammar. *JPh* **18**. 321–334.
Kiparsky, Paul (1985). Some consequences of Lexical Phonology. *Phonology Yearbook* **2**. 85–138.
Kirchner, Robert (1999). Preliminary thoughts on 'phonologization' within an exemplar based speech processing model. *UCLA Working Papers in Phonology* **2**. 207–231.
Kong, Qing-Ming (1987). Influence of tones upon vowel duration in Cantonese. *Language and Speech* **30**. 387–399.
Kruschke, John K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review* **99**. 22–44.
Labov, William (1971). Methodology. In William Orr Dingwall (ed.) *A survey of linguistic science*. College Park: University of Maryland Linguistics Program. 412–497.
Labov, William (1975). On the use of the present to explain the past. In Luigi Heilmann (ed.) *Proceedings of the 11th International Congress of Linguists*. Vol. 2. Bologna: Mulino. 825–851.
Labov, William (1994). *Principles of linguistic change*. Vol. 1: *Internal factors*. Oxford & Cambridge, Mass.: Blackwell.
Labov, William, Mark Karen & Corey Miller (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* **3**. 33–74.

Labov, William, Malcah Yaeger & Richard Steiner (1972). *A quantitative study of sound change in progress*. Philadelphia: U.S. Regional Survey.

Ladefoged, Peter (2000). *Vowels and consonants: an introduction to the sounds of languages*. Malden, Mass. & Oxford: Blackwell.

Lee, Wai Sum (1999). *A phonetic study of the speech of the Cantonese-speaking children in Hong Kong*. PhD dissertation, City University of Hong Kong.

Manaster Ramer, Alexis (1996). A letter from an incompletely neutral phonologist. *JPh* **24**. 477–489.

Milroy, James & John Harris (1980). When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World-Wide* **1**. 199–210.

Nosofsky, Robert M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology : General* **115**. 39–57.

Nunberg, Geoffrey (1980). A falsely reported merger in eighteenth-century English: a study in diachronic variation. In William Labov (ed.) *Locating language in time and space*. New York: Academic Press. 221–250.

Ohala, John (1993). The phonetics of sound change. In Charles Jones (ed.) *Historical linguistics: problems and perspectives*. London & New York: Longman. 237–278.

Pierrehumbert, Janet B. (1990). Phonological and phonetic representation. *JPh* **18**. 375–394.

Pierrehumbert, Janet B. (2001). Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (ed.) *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: Benjamins. 137–157.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology* 7. Berlin & New York: Mouton de Gruyter. 101–139.

Port, Robert F. & Penny Crawford (1989). Incomplete neutralization and pragmatics in German. *JPh* **17**. 257–282.

Port, Robert F. & Michael L. O'Dell (1985). Neutralization of syllable-final voicing in German. *JPh* **13**. 455–471.

Sancier, Michele L. & Carol A. Fowler (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *JPh* **25**. 421–436.

Sankoff, Gillian (2004). Adolescents, young adults and the critical period: two case studies from 'Seven Up'. In Carmen Fought (ed.) *Sociolinguistic variation : critical reflections*. Oxford: Oxford University Press. 121–139.

Tenpenny, Patricia L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review* **2**. 339–363.

Tieszen, Bozena (1997). *Final stop devoicing in Polish : an acoustic and historical account for incomplete neutralization*. PhD dissertation, University of Wisconsin, Madison.

Trudgill, Peter (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Vance, Timothy J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica* **33**. 368–392.

Vance, Timothy J. (1977). Tonal distinctions in Cantonese. *Phonetica* **34**. 93–107.

Vance, Timothy J. & Carol A. Walker (1976). Tone and intonation in Cantonese. *CLS* **12**. 640–655.

Wang, Li (1932). *Une prononciation Chinoise de Po-Pei* (*Province de Kouang-si*) *étudiée à l'aide de la phonétique expérimentale*. Paris: Libraire Ernest Leroux.

Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *JPh* **32**. 251–276.

Wedel, Andrew (2007). Feedback and regularity in the lexicon. *Phonology* **24**. 147–185 (this issue).

Whitaker, Katherine Po Kan (1955–56). A study on the modified tones in spoken Cantonese. *Asia Major* **5**. 9–36, 184–207.

Wong, Maurice Kuen-Shing (1982). *Tone change in Cantonese*. PhD dissertation, University of Illinois at Urbana-Champaign.

Wong, Patrick C. M. (1999). The effect of downdrift in the production and perception of Cantonese level tones. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: University of California. 2395–2397.

Wong, Patrick C. M. & Randy L. Diehl (1998). Effects of speaking fundamental frequency on the normalization of Cantonese level tones. *JASA* **104**. 1834.

Wong, Patrick C. M. & Randy L. Diehl (1999). The effect of duration on the perception of Cantonese level tones. *JASA* **106**. 2150.

Wong, Wai Yi Peggy (1996). *Tempo, processing rate and clarity drive in Hong Kong Cantonese connected speech*. MA dissertation, Hong Kong Polytechnic University.

Wong, Wai Yi Peggy (2004). Syllable fusion and speech rate in Hong Kong Cantonese. In Bernard Bel & Isabelle Marlien (eds.) *Actes des XXVes Journées d'Étude sur la Parole*. Fès, Morocco: AFCP. 255–258.

Yip, Moira (1980). *The tonal phonology of Chinese*. PhD dissertation, MIT.

Yip, Moira (1989). Contour tones. *Phonology* **6**. 149–174.

Yip, Moira (2002). *Tone*. Cambridge: Cambridge University Press.

Yu, Alan C. L. (2004). Explaining final obstruent voicing in Lezgian: phonetics and history. *Lg* **80**. 73–97.

Yuan, Jiahua (1983). *Hanyu fangyan gaiyao*. [*Survey of Chinese dialects*.] Beijing: Wenzi Gaige Chubanshe.