

1. BACKGROUND

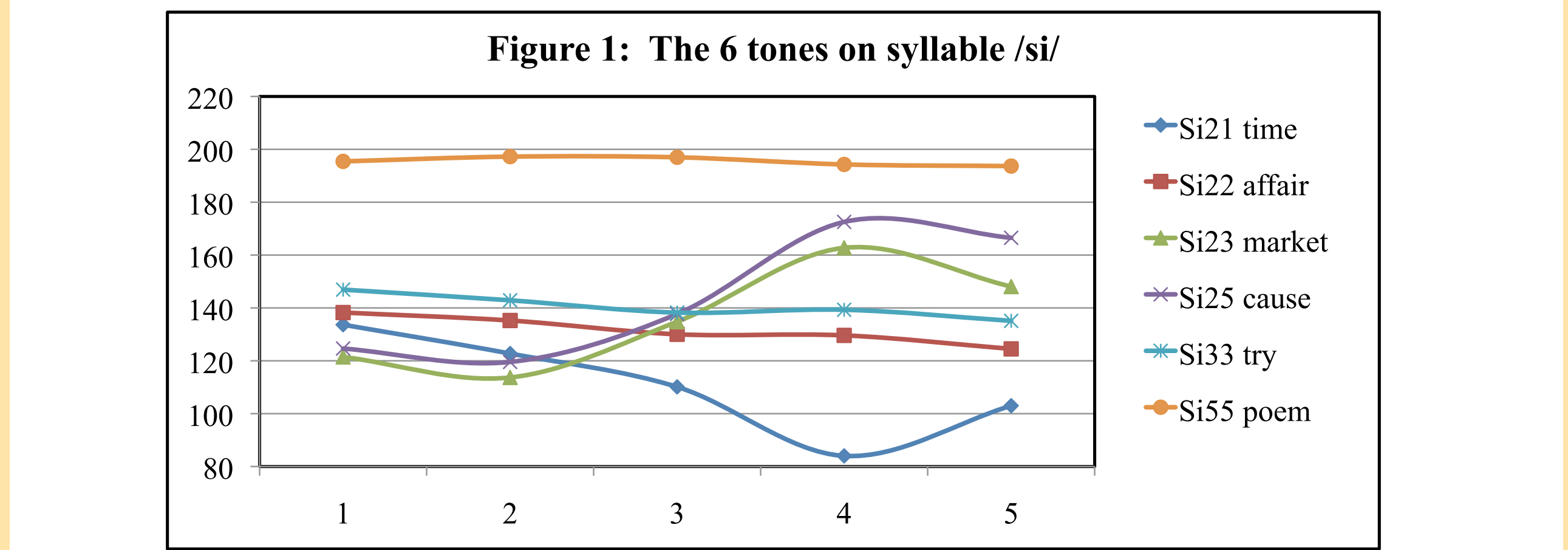
Northern and Southern Vietnamese tones seem to be perceptually organized along matrices in which perceptual cues are used by several tones (Brunelle 2009). Here is an example from Northern Vietnamese:

Table 1. Perceptual cues used by Northern Vietnamese listeners in colloquial speech		
	Modal	Glottalized or creaky
Low offset	huyền	hỏi
Rising	sắc	ngã
Else	ngang	nặng

Is this matrix shape just a coincidence or is it found in other East Asian languages? Cantonese is a good candidate because, like Vietnamese, it has a large tone inventory in open syllables.

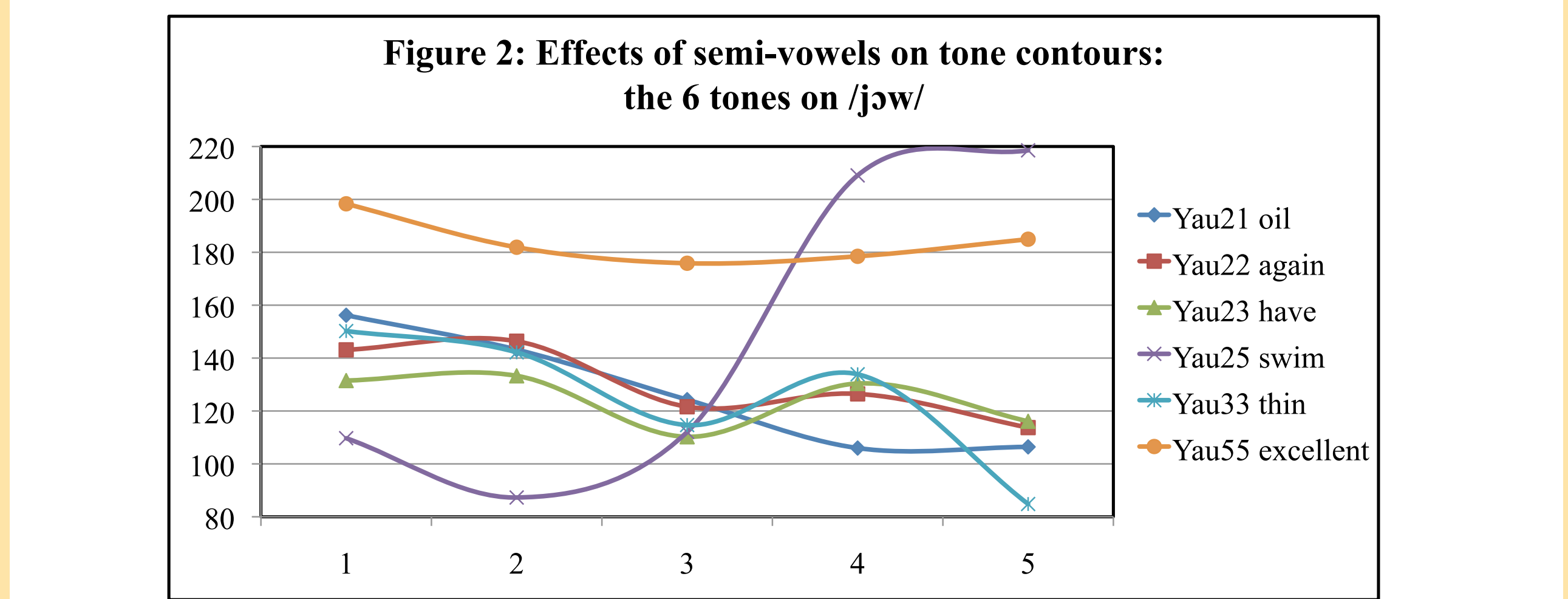
Cantonese has a large number of tones (6 on open syllables). Hence a crowded tone space:

- Three level tones: 22, 33 and 55
- Two rising tones: 23 and 25
- Four tones in the lower pitch range: 21, 22, 23, 33



Unsurprisingly, f0 is affected by segments.

- In Figure 2, a drop in f0 in the mid-part of most tones can be noticed. It is due to the lower intrinsic f0 in /ɔ/ than in semi-vowels.
- The perturbations due to intrinsic f0 make tones 22, 23, 21 and 25 very similar.



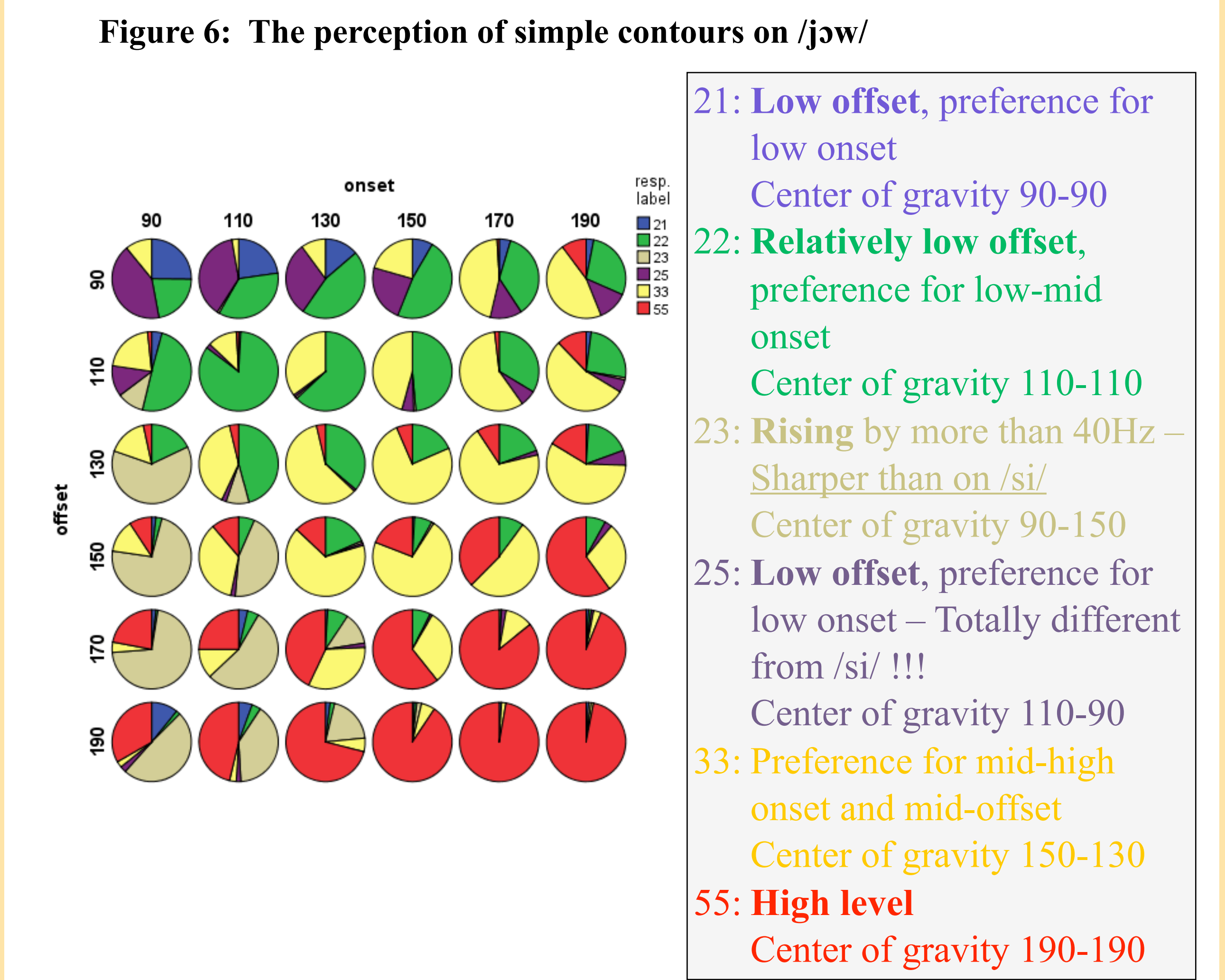
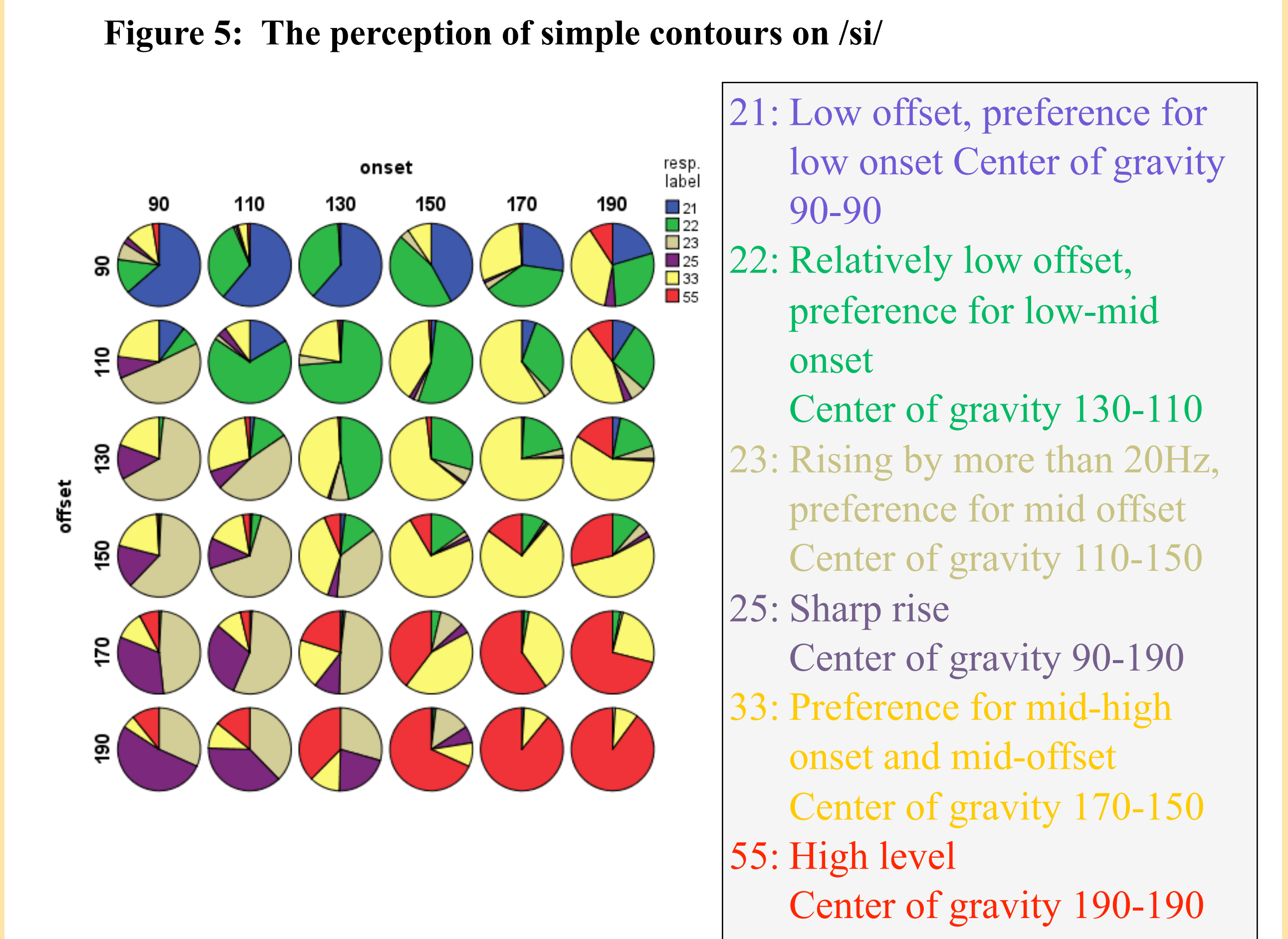
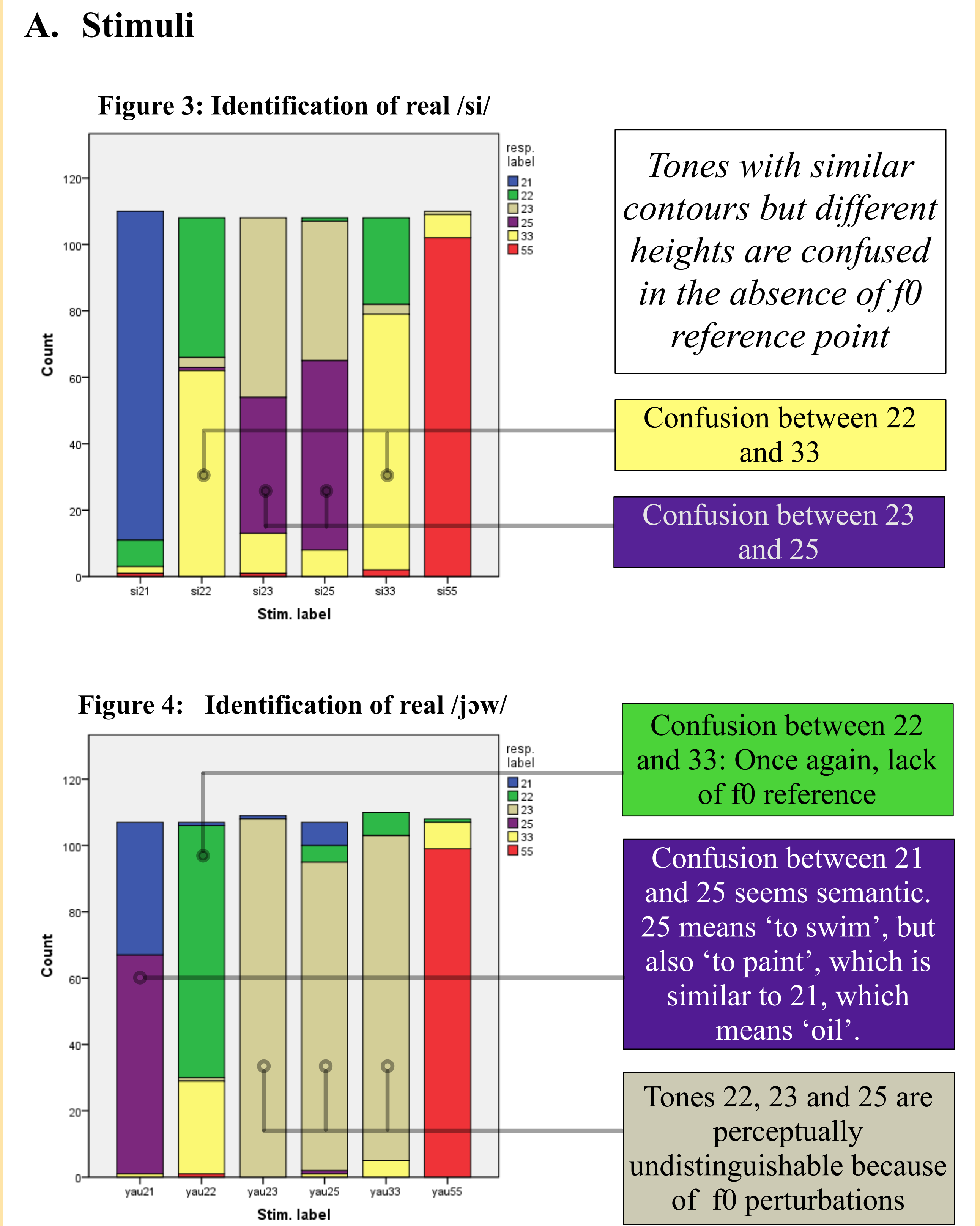
2. RESEARCH QUESTIONS

- How much confusion is there between tones?**
 - Quite a bit, even with natural stimuli (Cutler and Chen 1997; Khouw and Ciocca 2007)
 - Especially in the absence of reference point (stimuli presented in isolation) (Francis et al. 2003; Wong and Diehl 2003; Francis et al. 2006)
- What are the cues used for identification?**
 - Unitary contour (Francis et al. 2003; Li and Lee 2007, 2008) vs. specific points in the tone curve (Khouw and Ciocca 2007)
 - Are these cues redundant? (Brunelle 2009)
- How are similar tones distinguished?**
- How does intrinsic f0 and its effect on f0 affect perception?**
 - Do listeners compensate or do they get more confused?

3. METHODS

- Stimuli**
 - 12 natural words (6 /si/, 6 /jɔw/)
 - 72 stimuli (36 /si/, 36 /jɔw/) with simple re-synthesized f0 contours (6 onset targets X 6 offset targets)
 - 48 stimuli (24 /si/, 24 /jɔw/) with complex re-synthesized f0 contours (3 onsets X 3 midpoints X 3 offsets, minus level contours) Not reported here for lack of space, but feel free to ask!
- Subjects**
 - 10 subjects (5 males and 5 females)
 - All native speakers of Cantonese using it on a daily basis

4. RESULTS



5. DISCUSSION AND CONCLUSION

- There is a large amount of confusion even with natural stimuli**
 - Because of a lack of reference points for gauging f0 (Ref)
 - Greater confusion with /jɔw/ than /si/: Inability to filter effect of intrinsic f0?
- Cues used for identification of synthetic stimuli**
 - Rising vs. Non-Rising. Falls, on the other hand, are not perceptual cues.
 - Offset: A low offset is a necessary perceptual cue of tones 21 and 22. Tone 55 almost always require a high offset.
 - Tone onsets provide additional cues, but do not seem to define tonal identity. For example, tones with mid-offset are more identified as 22 if they have a low onset and as 33 if they have a high onset, but this effect is non-categorical.
- Differences in the perception of synthetic contours on /si/ and /jɔw/**
 - Rises**
 - A sharper rise must be present for 23 to be perceived on /jɔw/ than on /si/.
 - Rising contours are never perceived as 25 on /jɔw/.
 - This is probably due to the confusing effect of the intrinsic f0 of semi-vowels. Due to this added uncertainty, a sharper slope is necessary for the identification of rising tones 23 and 25.
 - The subjects realize that no instance of /jɔw/ sounds like tone 25 and therefore start identifying low level tones as 25. The reason is that tone /jɔw/ 21 means 'oil' which is semantically close to /jɔw/ 21 25, which means 'to paint' (Lexical effect discussed in Zhao and Jurafsky 2007).
 - Tone height**
 - Finally, the perceptual thresholds of several tones are lower on /jɔw/ than on /si/. This is an indication that listeners compensate for intrinsic f0 to a certain extent, even if it creates confusion. They subconsciously know that the nucleus of /jɔw/ has an inherently lower f0 than the nucleus of /si/.

D. Tone representation

A gross oversimplification of the perceptual cues used for tone identification would look like this:

	Offset			
Contour		Low	Mid	High
Rising			23	25
Non-Rising		21	22 (Low onset) 33 (High onset)	55

It is clear that the type of contour (rising vs. non-rising) and the height of offsets are used as perceptual cues. Interestingly, all cues are used in a redundant manner in such a way as to build a matrix where each box is occupied. However, the onset cue needed to distinguish tones 22 and 33 introduces a dimension that is not fully exploited in the matrix. Therefore, the perceptual space of Cantonese is less tightly packed than that of Northern Vietnamese (Brunelle 2009).

Further, contrary to what the table suggests, there are no categorical cutoffs. This suggests that reinterpretation of the perceptual cues into tone letters would be too hasty.

7. REFERENCES

Brunelle, Marc. 2009. Tone Perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37 (1):79-96.

Cutler, E. A., and H. C. Chen. 1997. Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics* 59 (2):165-179.

Francis, A. L., V. Ciocca, and B.K. C. Ng. 2003. On the (non)categorical perception of lexical tones. *Perception and Psychophysics* 65 (7):1029-1044.

Francis, A. L., V. Ciocca, N. K. Wong, W. H. Leung, and P. C. Chu. 2006. Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America* 119 (3):1712.

Khouw, Edward, and Valter Ciocca. 2007. Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35:104-117.

Li, Yujia, and Tan Lee. 2007. Perceptual Equivalence of Approximated Cantonese Tone Contours. In *Interspeech*. Antwerp, Belgium.

Li, Yujia, and Tan Lee. 2008. A Perceptual Study of Approximated Cantonese Tone Contours. Paper read at 6th International Symposium on Chinese Spoken Language Processing, at Kunming, China.

Wong, Patrick C. M., and Randy L. Diehl. 2003. Perceptual Normalization for Inter- and Intra-talker Variation in Cantonese Level Tones. *Journal of Speech, Language, and Hearing Research* 46:413-421.

Zhao, Yuan, and Dan Jurafsky. 2007. The effect of lexical frequency on tone production. Paper read at 16th International Congress of Phonetic Sciences, at Saarbrücken.

Zsiga, Elizabeth, and Rattima Nitisaroj. 2007. Tone Features, Tone Perception, and Peak Alignment in Thai. *Language and Speech* 50 (3):343-383.