Pitch Contour Shape Matters in Memory

Amelia E. Kimball, Jennifer Cole

University of Illinois at Urbana Champaign, USA

akimbal2@illinois.edu, jscole@illinois.edu

Abstract

The Autosegmental-metrical model of prosody [1,2] holds that pitch melodies can be modeled with level low and high tones; information about the shape of the pitch contour is not part of the phonological representation. Yet recent results [3,4] show that contour shape affects the perception of tone height and timing. A pitch plateau that maintains a level pitch at its peak will be perceived as higher and/or having a later accent than a sharp peak of the same height. In this study we ask whether contour shape is encoded in the mental representation of pitch accent by testing memory for the H* pitch accent of American English, realized as a peak or plateau. We establish that, as predicted by recent research, pitch shape affects perception. Then we test these same distinctions in a memory task. Our findings show that pitch plateaus are better discriminated than peaks, and that this advantage grows larger when memory load is higher. We argue that this shows contour shape matters, not just psychoacoustically in immediate perception, but also in memory, and that shape may therefore be posited to be included in the phonological representation of pitch accent.

Index Terms: prosody, pitch accent, pitch contour, autosegmental metrical model, episodic memory, memory for speech, exemplar models, abstractionist models.

1. Introduction

In English, intonation is linguistically meaningful, conveying information about the structural context of a word and its information status along with paralinguistic information about the speaker's affect and emotional state. Like any other linguistic feature, intonational features marking pitch accent or phrasal boundary are variable from speaker to speaker and instance to instance. In the face of this variability, the central question for speech perception and language comprehension research becomes: How do listeners perceive prosodic features given the extent of variability in the signal, and what aspects of phonetic detail, if any, are listeners sensitive to? Under abstractionist theories of intonation such as the Autosegmental-Metrical (AM) theory, listeners are expected to adapt to this variability by mapping perceived utterances onto a sequence of one or more abstract intonational features, which are categorically and meaningfully distinct. Exemplar theories of speech perception and memory present an opposing view, in which listeners create a phonetically detailed representation of each heard instance, and where categorical pitch accent and boundary features emerge from the statistical distribution of acoustic parameters over the stored instances [5-7].

To better understand the nature of the mental representation of intonational features, we need to know what aspects of the phonetic detail of intonational features listeners encode in memory. Here we focus on one type of prosodic variability: pitch accent contour shape, and specifically peaked vs. plateau-shaped accents. Under AM theory, intonation contours are modeled as combinations of low and high level tone features, and the shape of the pitch contour is not specified in the phonological representation of pitch accent. For example, the two contours in figure 1 would both be marked as H* in ToBI transcription [8]. The distinction between a peaked shaped accent and a plateau shaped accent is not known to be contrastive in any language [3] and so by modeling both of these contours as H*, meaningfully contrastive detail is retained, while non-contrastive information about contour shape is lost.

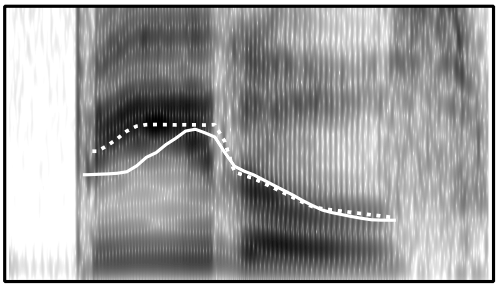

Figure 1: *Two H* manipulations of the word "beetles". The plateau contour (dashed line) and the peaked contour (solid line) were both resynthesized from the same utterance.*

Two areas of research, however, suggest that contour shape may in fact be important for comprehension of intonation. First, it is known that contour shape affects the perception of pitch height and timing, in that plateau shapes are heard as higher and/or later-timed than peak shapes with the same maximum value [3, 9-11]). Furthermore, while shape itself may not be contrastive, alignment and timing are contrastive in some languages, and so the difference in shape that leads to a perceptual difference in height and timing may affect the perception of speech category (see D'Imperio's work on Neapolitan Italian [11]).

Secondly, our previous work has shown that subcategorical phonetic detail for pitch accent is retained in memory [12]. In a memory task, listeners were able to remember not just the phonological category of a speech sample but also specific details of pitch and duration that are not known to be meaningful. We argue that these findings rule out the strictest interpretation of abstractionist theories, because if listeners map sounds to an abstract category, and remember only the category, sub-categorical detail will not be included in memory at all.

We test listeners' memory for heard pitch accents using the memory paradigm of our prior study [12]. We use discrimination tasks incorporating memory for two reasons.

Firstly, the use of memory builds on research in the sentence processing literature which shows that memory for meaningful linguistic information is privileged above memory for the specific form of the information. For example, when subjects are exposed to a sentence and then later asked if a test sentence is one they read or heard before, they have difficulty rejecting sentences that have the same meaning as the sentences they heard, but with different word order [13-15] or synonyms [16]. This work indicates that information about form is quickly forgotten while semantic information is stored in long term memory. Due to these robust findings that memory is strongest for meaning rather than form, we expect that meaningful prosodic information will be remembered better, while details of prosodic form will be remembered less well. We test whether contour shape is perceived and represented in the grammar by testing whether after initial exposure, listeners can successfully reject test utterances that differ in contour shape from those that they heard before.

Secondly, we use a memory technique because it connects our specific research question with a broader question in speech perception: when listeners hear speech, do they encode and store in memory each phonetically detailed instance of a heard word or phrase, or only the abstract speech units (e.g., phonemes) that comprise the heard speech? Abstractionist views, including research cited above about memory for sentences as well as experiments showing the categorical perception of phonemes (e.g. [17-19]) suggest that sensory information fades quickly in memory and only meaning, or meaningful categories, remain. However, subsequent studies reveal that listeners are sensitive to within-category acoustic detail, and that such detail can influence phoneme and lexical identification [20,21]. A further finding is that phonetic detail is linearly encoded prior to categorization of the speech input, and maintained through late stages of perceptual processing [22] These findings are consistent with exemplar models, in which listeners represent phonetically detailed speech input as episodic memory traces, and build up statistical distributions of these traces rather than categories. Further support for exemplar models comes, for example, from findings that listeners are better able to recognize a sentence they have heard before when presented with the exact utterance, including voice information and background noise. [5,6] This suggests that even non-linguistic acoustic information is retained in memory, at least to some extent.

Against the background of this prior research, we ask whether contour shape is encoded in the phonological representation of pitch accent in American English. We test two competing predictions. Under abstractionist theories, both peaks and plateaus are modeled with a high tone, and so both are predicted to be remembered equally well. Under exemplar models, contour shape (and all other details) are encoded, and therefore different contours may not be remembered equally well.

2. Method

This experiment uses the memory paradigm of [12] where memory for pitch accent is tested through AX discrimination tasks. The AX task tests participants' ability to perceive intonational differences in two productions of the same word when presented in immediate succession, or with an intervening delay. The present experiment focuses on listeners' ability to perceive and remember differences in pitch accent contour shape. The experiment is a two by two design: we test two contour shapes (peaks vs. plateaus) in two tasks (AX vs. delayed AX). If both contour shapes are remembered equally easily, we expect memory performance for the two contours to be similar across the two tasks.

2.1. Stimuli

Stimuli were words excised from natural utterances, created by a trained linguist who was not part of the research team. All words were content words produced with an H* which were designed to have all voiced segments (e.g. "movies", "beavers", "vegans", see also figure 1). For each natural utterance, four resynthesized versions were created: a lowered peak, a raised peak, a lowered plateau, and a raised plateau. Contour shapes were created by stylizing the pitch contour to 10 Hz (that is, smoothing the contour to a step size of 10 Hz) using Praat [23] and manually adjusting the contour into either a peak shape or a 75ms plateau following [3], keeping the starting and ending points of the pitch rise the same. Next, these peaks and plateaus were shifted up 25 Hz or down 25 Hz, as in [12]. Only resynthesized stimuli that had been shifted up or down were included in the experiment, so subjects were never asked to distinguish between a naturally produced token and a resynthesized token.

2.2. Participants

120 participants were recruited through Amazon Mechanical Turk, 30 in each of the four experiments. All participants were located in the United States and were self-reported native English speakers with no self-reported hearing problems.

2.3. Procedure

All experiments were conducted online using Amazon Mechanical Turk. In experiments 1 and 2, participants heard two versions of the same word with one second of intervening silence. They were immediately asked to click on a button to indicate if they were "the exact same recording or different recordings." (an AX task) The pairs of words were either the same recording (1/2 of the trials), or they differed in peak height (experiment 1) or plateau height (experiment2).

Experiments 3 and 4 use the same stimuli but add a delay and interference, to make discrimination more difficult. Listeners heard four different words produced by the same speaker (exposure), then a tone, and then another presentation of a word from the exposure phase (test). They were asked to report whether the test word was "exactly the same recording" as the exposure version. Again, the pairs of words were either the same recording (1/2 of the trials), or they differed in peak height (experiment 3) or plateau height (experiment4).

Each participant took part in only one experiment, after which they completed a post-test AX task which tested their perception of pure tones. 36 pairs of pure tones that differed by 25 Hz (half of the magnitude of the discriminations above) and varied from 75-300 Hz were used. Participants heard a tone, then a second of silence, and then a second tone that was either the same (half of the trials) or differed by 25 Hz (half of the trials). Participants were scored for discrimination accuracy, and the score on the post test was used as an inclusion criterion for the results reported in section 3. Those who scored 2 standard deviations below the mean, or lower, were not included in the analysis presented here (5 out of 120 subjects).

Since the post test is less difficult than the experiment itself it was included as a way to ensure that subjects were listening carefully and not guessing, even if their scores were near chance in experiments 1, 2, 3 or 4. Subjects who did well at the post test but were near chance at the other experimental task were included in our results.

3. Results

Each response was coded as correct (either a hit or a correct rejection) or incorrect (either a miss or a false alarm). Accuracy was analyzed in a mixed effects logistic regression using the lme4 software package in R [24] Fixed effects included task (Delay vs. AX), and contour (plateau vs. peak), and the interaction between the two. The random effects structure was determined by backwards model selection starting with the maximal random effects structure justified by the design. [25] The maximally converging random effects structure included random intercepts for subject and item, and random slopes for subject.

Our three main findings were that accuracy was significantly lower in the delay task (Estimate: -0.50, SE: 0.047, p<.001), accuracy was higher in plateau contours than in peak contours (Estimate: 0.76, SE: 0.028, p<.001), and there was a significant interaction of task by contour shape (Estimate: 0.66, SE: 0.056, p<.001). Overall, results of our model indicate that listeners are better at discriminating pitch differences in contours with a plateau shape, rather than a peak shape, and better at the AX task than the Delayed AX task, even taking into account differences in subjects and items. Furthermore, this plateau advantage is shown even more strongly in the delay task, which taxes memory.

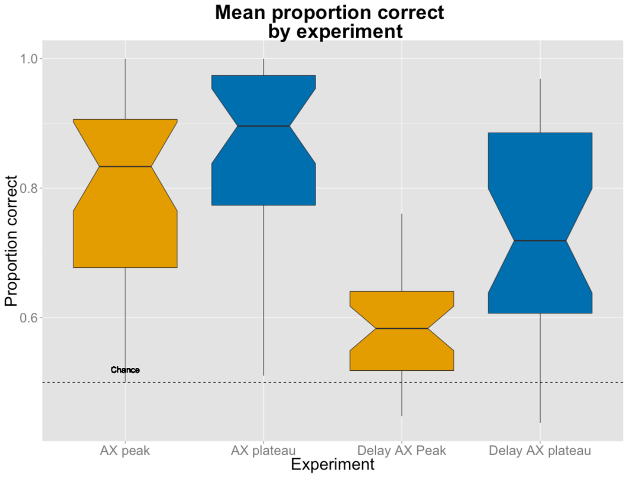

Figure 2: *accuracy score presented by experiment. Peak contours are in orange, while plateau are in blue.*

Results of the post test showed that even those participants who were close to chance in experiments 1,2,3, or 4 were attending to files and listening carefully. A linear regression showed that across all experiments score on the post test was a significant predictor of score in the experiment (*p*=.027), but it explained very little of the variance (multiple r^2=.0498)

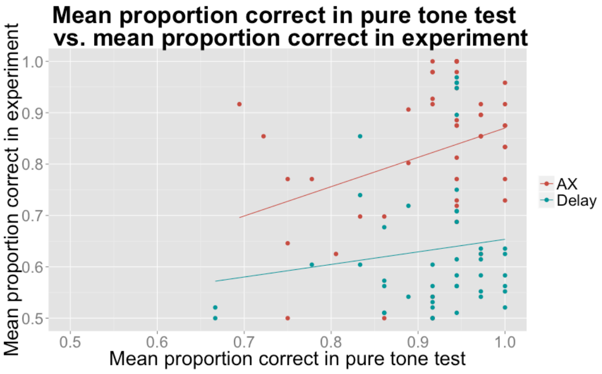

Figure 3: *Post test score vs. experiment score*

4. Discussion

The purpose of this study was to test whether contour shape is included in the phonological representation of pitch accent. We tested two contour shapes which are modeled with a high tone in the AM framework (H* in ToBI notation). If these contours are indeed mapped to the same phonological representation, we predicted that they will be discriminated at the same rate, both in a perceptual task and a memory task. Instead, our study had two main findings: plateaus were discriminated slightly better after a short time lag, and this advantage increased with higher memory load. Our results suggest that listeners are sensitive to subcategorical differences in contour shapes, and this sensitivity translates to a difference in memory.

We attribute the observed advantage for plateau contours to two factors. Firstly, the main effect of contour shape is no doubt driven by the psychoacoustic salience of these contours—plateau contours involve a longer time at a pitch maximum, giving listeners a larger target to hear. Given this, we would expect the observed across-the-board advantage for plateaus. However, if our results were solely due to a perceptual advantage of plateaus over peaks, we would expect to find the same effect, with the same magnitude, in the harder delayed AX task. Instead, our second main finding is that the observed advantage in the AX task was *larger* in the delayed AX task, indicating that the plateau advantage is greater in memory.

The second factor which may drive the plateau advantage is the relative frequency of plateau shapes in speech, when compared to peak shapes. Plateaus are more common in everyday speech, [3] and indeed the vocal tract is not capable of changing frequency quickly enough to make very sharp peaks. If peaks are rare, they may not constitute good members of the H* category, or may be outliers.

Ultimately, this plateau advantage is consistent with three different theoretical positions. Firstly, because plateau contours are more common, our results are consistent with an exemplar model, in which each instance of prosody is remembered faintly and builds up a cloud of episodic memory traces [5]. If plateaus are more common, then plateau contours will conform more closely to the exemplar-based representation built up over time. This is consistent with a growing body of work in sentence processing that suggests that listeners are sensitive to the statistical distribution of language features over time. [26-28]

Additionally, our results are also consistent with an enriched abstractionist view, wherein contour shape is specified in the phonological representation. If the abstract category of H* is specified for shape, then plateau contours which map to the H* category will be successfully mapped, and better remembered, as they are in our study

Lastly, under an abstractionist theory wherein pitch contours are modeled as level tones, such as AM, it could be that peak contours are simply not typical intonation contours of English, and therefore are not reliably mapped onto the H* accent category and encoded into memory.

Crucially, the thread which holds together all of these possibilities is that in all cases *contour shape matters.* We argue that our results show that variability in contour shape is not simply noise to be abstracted away from, but rather important information that may affect perception (as shown by [10] and [3]) and is encoded in memory. We echo Barnes and others in calling for enrichment of phonological representations of prosody, as in Barnes et al.'s Tonal Center of Gravity, [29] or moving still further toward exemplar-based approaches.

5. Conclusion

Overall, our results show an advantage for plateau contours in memory. Based on these results, we argue against strictly abstractionist views of memory for speech, and argue that information about contour shape is part of the phonological representation of pitch accent.

6. References

[1] J. Pierrehumbert, *The Phonetics and Phonology of English Intonation*. Ph.D. Dissertation, MIT. 1980.

[2] D. R. Ladd, *Intonational Phonology*. 2nd ed. Cambridge: Cambridge University Press. 2008.

[3] J.Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux "On the nature of perceptual differences between accentual peaks and pleateaux" in Oliver Niebuhr &Harmut Pfitzinger (eds.), *Prosodies: Context, Function, Communication*. Berlin/New York:Mouton de Gruyter, 2012.

[4] F. Cangemi, *Prosodic Detail in Neapolitan Italian*. Berlin: Language Science Press ,2014.

[5] S. D. Goldinger, "Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory," *Journal of Experimental Psychology: Learning*, vol. 22 no. 5, pp.1166-1183, 1996.

[6] A. Pufahl and A. G. Samuel, "How lexical is the lexicon? Evidence for integrated auditory memory representations," *Cognitive Psychology,* vol. 70, pp. 1-30, 2014.

[7] J. Bybee, "From usage to grammar: The mind's response to repetition" *Language,* vol. 82, no. 4, 711-733

[8] K.E. Silverman, et al. "TOBI: a standard for labeling English Prosody" *The Second international conference on Spoken Language Processing, ICSLP,* Banff, Alberta, Canada, October 13-16 1992.

[9] J. 't Hart "F0 stylization in speech: straight lines versus parabolas." *Journal of the Acoustical Society of America* vol. 90, no. 6, pp. 3368-3370, 1991.

[10] O. Niebuhr. "The Signaling of German Rising-Falling Intonation Categories – The Interplay of Synchronization, Shape, and Height." *Phonetica,* vol. 64, no. 2-3, pp. 174-193,2007.

[11] M. D'Imperio*The Role of Perception in Defining Tonal Targets and their Alignment.* Ph.D. Dissertation, The Ohio State University. 2000.

[12] A.E. Kimball, J. Cole, G. Dell, S. Shattuck-Hufnagel, "Categorical vs. Episodic Memory for Pitch Accents in English", *proceedings of the International Congress of Phonetic Sciences*. Glasgow, UK, 2015.

[13] J. S. Sachs, "Recognition memory for syntactic and semantic aspects of connected discourse" *Perception and Psychophysics,* vol. 2, pp. 437-442, 1967

[14] J.R. Anderson, "Verbatim and propositional representation of sentences in immediate and long-term memory" *Journal of Verbal Learning and Verbal Behavior*, vol. 13, no.2, pp.149-162,1974.

[15] M.C. Potter and L. Lombardi, "Regeneration in the short-term recall of sentences" *Journal of Memory and Language*, vol. 29, no. 6, pp. 633-654, 1990.

[16] W.F. Brewer, "Memory for ideas: Synonym substitution" *Memory and Cognition*, vol. 3, no. 4, pp. 458-464, 1975.

[17] A.S. Abramson and L. Lisker, "Discriminability along the voicing continuum: Cross-language tests." *Proceedings of the sixth international congress of phonetic sciences*. vol. 196, no. 7. Academia Prague, 1970.

[18] Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

[19] Repp, B.H. (1984). Categorical perception: Issues, methods and findings. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 244–335). New York, NY: Academic Press.

[20] McMurray, B., Tanenhaus, M.K., & Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*, B33–B42.

[21] McMurray, B., Aslin, R.N., Tanenhaus, M.K., Spivey, M., & Subik, D. (2008). Gradient sensitivity to within-category variation in speech: Implications for categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 34, 1609–1631.*

[22] Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). "Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech." *Psychological Science*, 21(10), 1532-1540.

[23] P. Boersma and D. Weenink "Praat: doing phonetics by computer" [Computer program]. Version 6.0.05, retrieved 8 November 2015 from http://www.praat.org/

[24] D. Bates, M.Mächler, B.Bolker, and S. Walker. "Fitting linear mixed-effects models using lme4". *arXiv preprint arXiv:1406.5823*. 2015.

[25] D.J.Barr, R. Levy, C. Scheepers, and H.J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal." *Journal of Memory and Language*, vol.68, no. 3, pp. 255-27, 2013.

[26] M.C. MacDonald, "Distributional Information in Language Comprehension, Production, and Acquisition: Three Puzzles and a Moral" in Brian MacWhinney (ed.) *The Emergence of Language*, pp.177-196. Mahwah, NJ, USA:Erlbaum. 1999.

[27] F. Chang, G.S. Dell, K. Bock, "Becoming Syntactic" *Psychological Review,* vol. 113, no.2, pp.234-272, 2006.

[28] D. Kleinschmidt and F. Jaegar, "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel" *Psychological Review* vol. 122, no. 2, pp. 148-203,2015.

[29] J. Barnes, N. Veilleux, A. Brugos, S. Shattuck-Hufnagel, "Tonal Center of Gravity: A global approach to tonal

implementation in a level-based intonational phonology" *Laboratory Phonology*, vol. 3, no. 2, pp.337-383, 2012.