The Auditory Kappa Effect in a Speech Context

Alejna Brugos, Jonathan Barnes

Boston University, Boston, Massachusetts, USA

abrugos@bu.edu, jabarnes@bu.edu

Abstract

Results of a perception experiment show that the size of the pitch difference between spoken words on either side of a pause affects perception of pause duration, demonstrating that the auditory kappa effect can obtain in speech materials. The auditory kappa effect is an illusion whereby, in a sequence of tones and intervening silent intervals, the perception of the duration of the silent intervals is influenced by the relative pitch of the tones, and has previously been demonstrated only using musical or simple (sinusoidal) tones. This study extends these findings to spoken language materials, suggesting that the influence of pitch on the perception of speech timing needs to be further examined. The results of this study, along with studies showing the complex effects of dynamic pitch on perceived duration, are discussed with respect to implications for phonetic measures of prosodic events, such as relative boundary size and syllable lengthening.

Index Terms: speech prosody, duration perception, auditory kappa effect, pitch, timing

1. Introduction

Timing aspects of speech prosody have generally been investigated using objective measures of time, such as duration in ms. of segments, syllables, or silent intervals. Likewise, aspects relating to pitch are typically investigated using local measures of f0 at specific events, with little attention given to the temporal context in which they occur. However, there are suggestions from the literature that the perception of pitch and timing are not entirely independent.

Several studies (Yu, 2010 [1], Cumming, 2011 [2], *inter alia*) have found evidence that listeners perceive vowels with dynamic f0 as longer than those with static f0. Similarly, research in the perception of (non-speech) tone glides has shown that the velocity of pitch change in the glide (relative to standards with differing pitch velocity) influences listeners' perception of the duration of the glide, and vice versa (Henry 2011 [3]). A further body of research examines the mutual effects of timing and pitch perception in sequences of level tones and silent intervals, with two related phenomena known as the auditory tau and kappa effects. The auditory tau effect refers to the influence of temporal proximity on listeners' perception of pitch, and the kappa effect conversely describes the influence of pitch proximity on time perception. This paper will focus on the auditory kappa effect, and will examine whether it obtains in speech materials as it does in simple tone perception.

1.1. The kappa effect: time and space

The kappa effect refers most generally to a phenomenon whereby the perception of time intervals is distorted based on the perception of events in the spatial dimension. Specifically, the distribution in space of a sequence of three events can affect the perceiver's estimate of the time between each of the three events. The kappa effect has been demonstrated in diverse perceptual domains, involving distances between objects or events presented visually (Cohen *et al.*, 1953 [4]), auditorily (Sarrazin *et al*. 2005 [5]), and haptically as well (Grondin *et al*., 2011 [6]). In all these cases, variation in the distances between lights, sounds, or touches on the skin presented to subjects in succession has led to systematic differences in the perception of the time intervals separating them as well. Under certain conditions, individuals perceive stimuli that occur farther apart in space (e.g., flashes of light, or bursts of sound) as occurring farther apart in time than those occurring closer together spatially. In short, it would seem that the mind expects that a greater distance will take longer to traverse than a shorter distance, and adjusts perception accordingly.

1.2. The kappa effect in auditory perception: time and the frequency domain

In the domain of auditory perception, a primary direction of research has tested whether the kappa effect can be observed not just in the influence on temporal perception of distances across intervals in physical space, but of distances across analogous intervals in pitch space as well.

A body of research has suggested that the use of spatial metaphors (e.g., high and low) to describe the frequency domain is based on commonalities in the perception of space and frequency (see Henry 2011 [3] for an overview). Consistent with this assumption, Cohen *et al.* (1953 [4], 1954 [7]) hypothesized that in a sequence of tones, the frequency of the tones would distort the perception of time between tones, and called this predicted result the auditory kappa effect. While initial results were inconclusive, subsequent studies (Shigeno 1986 [8], 1993 [9], Crowder & Neath 1995 [10], MacKenzie 2007 [11] and Henry & McAuley 2009 [12], *inter alia*) found significant effects from pitch on timing perception, supporting the existence of the auditory kappa effect. Figure 1, below, illustrates this effect, whereby relative frequency affects timing perception. The tone sequences represented are each sets of 3 tones of equal duration, separated by 2 silent intervals (t1 and t2) also of equal duration. In the lefthand sequence, where the middle tone is closer to the first tone than to the third, subjects tend to perceive relative timing as shifted such that t1 is shorter than t2. Similarly, in the sequence on the right, the large change in pitch between the first and middle tones and the relative pitch proximity of the middle tone to the last influences subjects to indicate that t1 is longer than t2.

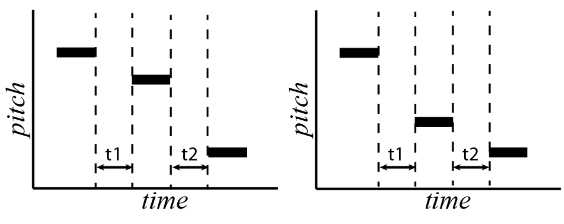

Figure 1*: A schematic example of the auditory kappa effect, whereby tone height affects timing perception. The silent intervals (t1 and t2) are of equal duration, but t1 is perceived as shorter at left, longer at right.*

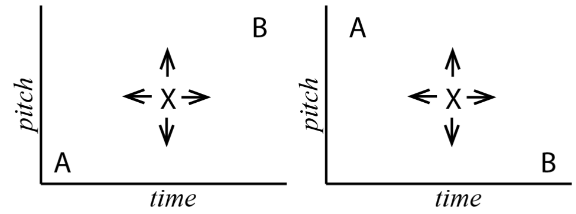

Figure 2: *A schematic of the kappa cell paradigm for testing the auditory kappa effect for two pitch change directions: ascending (left) and descending (right). In each frame, sound events A and B are fixed in pitch space, and in time relative to each other. Only the intermediate event X changes, and does so in both time and pitch space.*

Shigeno (1986 [8], 1993 [9]) devised a method for testing the kappa effect dubbed the kappa cell by MacKenzie (2007) [11]. A kappa cell consists of a three-tone sequence of the type AXB in which the A and B tones are fixed in f0 space and in time relative to each other, but both timing of inter-tone pauses and pitch of the X tone are systematically varied. (See Figure 2.) Subjects are asked to listen to the sequence, and judge whether the middle tone is closer in time to A or B.

Work testing for the possibility of an auditory kappa effect involving the influence of F0 patterns on the perception of speech timing has so far not appeared in the literature.[1] Given that objective measures of timing, including pause duration, final lengthening, etc., are regularly used to operationalize relative boundary size (e.g., Wightman *et al.* 1992 [13]), the potential for f0 interference with the perception of duration suggests that reliance on raw duration measures as reflections of prosodic structure may be missing important aspects of the perception of speech timing.

2. An experiment

To determine whether the kappa effect can play a role in the timing perception of speech material, we constructed an experiment based on the AXB kappa cell paradigm (Shigeno, 1986 [8], MacKenzie 2007 [11], see Figure 2, above). Subjects judged whether X was closer in time to A or B in 70 target stimuli varying the timing and pitch of X. This experiment is intended as a first step in the direction of determining what role, if any, the auditory kappa effect might have in the perception of speech timing.

2.1. Methods

The experimental design was modelled closely after that of Shigeno (1986 [8], 1993 [9]) and subsequent modifications by MacKenzie (2007) [11], using the AXB kappa cell paradigm. Whereas both Shigeno and MacKenzie used sequences of pure tones and silent intervals, this experiment uses sequences of a single spoken word, separated by silent intervals, as the sound stimuli. In order to move away from pure tone and music perception, and to decrease the likelihood of the stimulus items sounding like sung musical notes, this study used f0 manipulations on a word with a dynamic f0 contour typical of spoken language, the rise-fall.

[1] Shigeno (1986) did conduct a study on the kappa effect in speech; in that study, however, the relevant dimensions were F1 movement and vowel category, rather than F0 and timing cues.

2.1.1. Creation of stimuli

A single production of the word *one* was spoken in isolation by a female native speaker of American English in citation form (ToBI H* L-L%) in a moderately low range for that speaker. A monosyllable with relatively short duration was chosen to best offer comparison to results from MacKenzie (2007), who used tones of 200 ms. duration. The token of the word *one* chosen for resynthesis was 306 ms. That production was resynthesized using Praat such that the f0 formed a symmetrical rise-fall shape, with the low at 150 Hz. and the maximum at 200 Hz. (These values were chosen arbitrarily based on their similarity to those of the natural productions.) A series of new tokens was then resynthesized such that the same rise-fall shape was shifted upwards in 1 semitone steps, with the highest of the series 8 semitones higher than the base token. Accordingly, the highest of the series had a maximum f0 of 315 Hz., a height within the speaker's natural range (as seen in previous elicitations).

Kappa cells were constructed for 2 directions of pitch movement: ascending and descending. For the ascending condition, the base *one* was chosen as A, and the highest *one* as B, and for the descending condition, this order was reversed. Files were concatenated such that X was selected from among the 7 intermediate pitch steps of 1 semitone each. The silence between A and X (t1) and X and B (t2) ranged between 410 and 590 ms., in steps of 20 ms, such that the sum of t1 and t2 always equaled 1 second, giving 10 time steps for the location of X. In all, there were 70 unique stimuli (10 time x 7 pitch steps) for each of 2 orders.

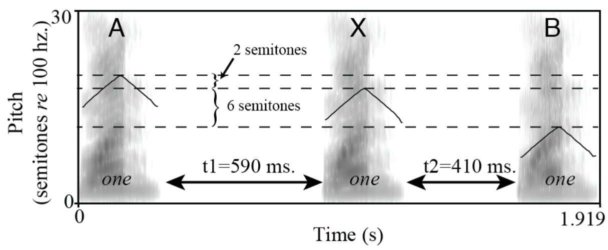

Figure 3: *A sample stimulus file, showing a sequence of the spoken word* one*, from the descending direction condition. The f0 contour of the first* one *(A) is 8 semitones above the f0 contour of the third* one *(B). The middle* one *(X) is set to 2 st. below A. The silent interval between A and X is set to 590 ms.*

2.1.2. Subjects

36 native speakers of American English, ranging in age from 18 to 26 years, participated in the study for a payment of $20. Subjects reported no hearing or speech deficits.

2.1.3. Presentation and task

Stimuli were presented in 4 blocks, each containing all 70 manipulations, with items randomized within each block, giving 280 trials. Subjects were assigned arbitrarily to one of 2 pitch change directions: the first 17 subjects were assigned ascending order, and the next 19 the descending order. The training and experimental sections took about 30 minutes, and subjects were given regular opportunities for breaks.

Subjects were seated in a quiet room facing a laptop, wearing headphones, and indicated responses using a Cedrus 530 response pad. Subjects were asked to indicate whether the middle *one* was closer in time to the first *one* or to the last. Subjects were explicitly instructed to base responses on temporal proximity alone, and likewise to do their best to ignore the changes in pitch.

2.1.4. Training and Screening

Before beginning the experiment, subjects read a brief written introduction to the study. They heard several recorded examples of sequences of words with varying intervening pauses while being presented with schematics showing differences in time and pitch level. To ensure that subjects were able to understand the task, and produce accurate responses based on timing, the experimental session included a brief training period using only the largest time differences (where t1 was equal to 410, 430, 570 and 590) and with only a subset of pitch steps. Subjects proceeded to the experimental phase when their cumulative correct score reached 75% (with at least 10 training trials presented), or upon completing 40 training trials. Most subjects (N=27) proceeded to the experimental section after 11 or 12 training trials.

To ensure that subjects were in fact able to perceive the timing differences at hand sufficiently well to provide a reliable reference for the additional effects of pitch on timing perception, a subset of the data collected, involving only the largest of the timing differences, was used to screen subjects for inclusion in the study. Responses for trials with the 2 smallest and the 2 largest t1 values (giving the 4 least ambiguous ratios of pause duration between t1 and t2) and for all 7 pitch steps (4 time steps x 7 pitch steps x 4 repetitions = 112 trials) were examined. Only subjects who correctly discriminated 75% or above of this subset were included. Only 4 subjects were excluded for not reaching this criterion. One additional subject was excluded due to equipment malfunction. The data from 31 subjects were included in the analysis (15 subjects from the ascending pitch condition, and 16 descending).

2.2. Results and analysis

Results presented here are from responses to 8680 experimental trials for 31 subjects. Figure 4 displays the proportion of responses that X is closer in time to B as a function of time step. Separate lines represent different pitch steps indicating the amount of shift (in number of semitones) from tone A. (Pitch steps have been coded such that the step size always indicates reference from tone A. This means that for the ascending order, step 1 is 1 semitone above the baseline, and for the descending order, step 1 is 1 semitone below the highest *one*.)

The general upward diagonal trend of the lines shows that subjects were responding primarily based on time: smaller pause durations between A and X (t1) result in fewer responses that X is closer to B, while larger t1 values result in more responses that X is closer to B. In addition, there appears to be an effect of pitch at all time values, as reflected in the separation of the pitch step lines. For steps when X is closer in pitch to A, there are proportionately fewer responses that X is closer in time to B. For pitch values that are closer in pitch to B (those that have the largest distance from the pitch of A), subjects responded in greater proportion that X is closer in time to B, even at the same T1 values.

We analyzed these results using mixed-effects logistic regression, implemented through the lme4 package (Bates & Maechler[14]) in R with response ("X closer to A" or "X

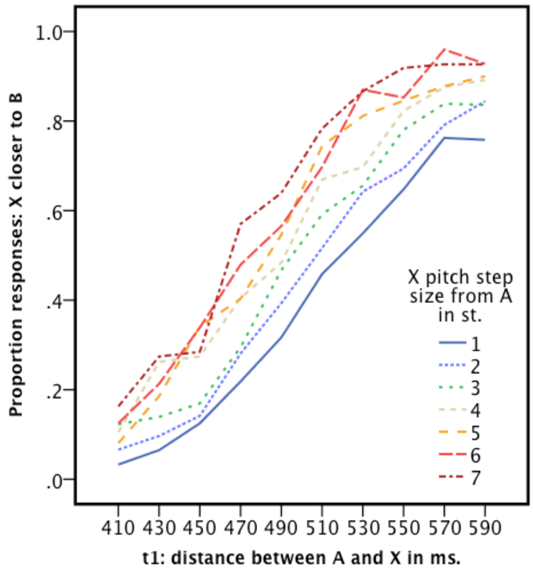

Figure 4: *Proportion of responses that X is closer in time to B for all 31 subjects as a function of time between A and X. Lines represent pitch steps between A and X in st. The separation of pitch lines reflects the effect of pitch on discrimination of timing: when X is closest in pitch to B (e.g., lines for 6 and 7 semitone steps), responses that X is closer in time to B are more frequent than for steps when X is closer to A (eg. steps 1 and 2) for each time step.*

closer to B") as the dependent variable, and time step, pitch step and direction of pitch change as fixed factors. Subject was included as a random effect (Baayen et al., 2008 [15]). The result was a model (N = 8533, log-likelihood = -4272) showing significant main effects of time step (Wald Z = 45.17, $p < .001$), pitch step (Wald Z = 10.24, $p < .001$) and direction (Wald Z = -4.08, $p < .001$), as well as a significant interaction between direction and pitch step (Wald Z = 2.96, $p < .001$).

3. Discussion

These results suggest that the kappa effect was indeed obtained in this experiment using speech materials. Much like in previous experiments using simple tones, the pitch changes across sequences of words with a fall-rise contour and intervening silent intervals led to systematic distortions in the perception of the duration of silent intervals.

The results of this study further suggest that the direction of pitch movement did play a role in the strength of the kappa effect in this study, with the effect being intensified in the descending order. The interaction of pitch step and direction reflects the greater influence in the descending condition of pitch on time perception for trials in which the pitch of X is closest to that of A (e.g., steps 1 and 2). (See Fig. 5, where the lines for pitch steps 1 and 2 appear shifted lower in the graph, as compared to the larger steps, e.g., 6 and 7, which stay high in both conditions.)

While MacKenzie (2007) [11] found no effect of pitch change direction, Henry & McAuley (2009) [12], who varied the velocity of the pitch change, found that the kappa effect could be intensified in downward pitch changes (relative to results for ascending pitch change) with slower pitch change velocities.

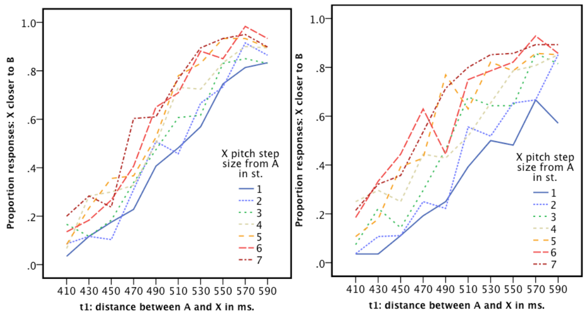

Figure 5: *Results separated by pitch change direction, displayed as in Fig. 4. The pitch step lines for the descending order (right) show greater separation than those of ascending (left).*

The variation in this result across studies could, therefore, be due to pitch velocity differences. As such, it may prove fruitful to vary both the length of pauses and the length of the spoken items (e.g., by using longer words or phrases) to explore the strength of the kappa effect in speech. It is also interesting to note that, while not invoking the kappa effect explicitly, Cumming (2011) [2] also found a difference in the magnitude of her results based on direction, in which descending tone movements showed greater effects than ascending movements on the perception of vowel duration.

While this study does show that the kappa effect can obtain in speech materials, the results, being based on a non-linguistic task, do not yet shed light on the extent to which the effect might transfer to meaning-based processing. A follow-up study is currently underway exploring how the same types of time and pitch variation affect judgments of phrasing or boundary strength, a topic that has typically been examined primarily in terms of domain-final or initial duration adjustments, and/or pause duration.

Exploration of the potential for mutual enhancement of pitch and timing cues in speech perception may shed light on a number of well-known phenomena in the literature on speech prosody. For example, building on pitch/timing effects discussed in Yu (2010) [1] and Cumming (2011) [2], work is currently planned to investigate the potential interaction of the dynamic f0 patterns associated with intonational boundary tones and durational patterns such as phrase-final lengthening. Likewise, connections might plausibly be drawn between the kappa effect demonstrated here, and the phenomenon of declination "reset" (e.g., Ladd 1988 [17], Fery & Truckenbrodt, 2005 [18]), whereby the relative scaling of certain cross-boundary pitch targets is said to play a role in signaling boundary strength as well. The possibility that all these cues enter into a complex set of trading relations will likewise be explored.

Such a direction of research may also shed light on perceived tone/timing mismatches in prosodic labelling, such as in the common case where the use of the 2 break index in ToBI is meant to signal clear tonal markers for a phrase boundary, in the absence of accompanying unambiguous duration cues. (Beckman & Ayers Elam, 1997 [16])

It should also be noted that this work thus far only concerns the effects of pitch on timing perception in very limited and isolated contexts in speech. MacKenzie (2007 [11]) found that the magnitude of the kappa effect could be modulated by preceding the kappa cell with a serial context: some contexts intensified the effect, and others nullified it. The extent to which the kappa effect holds in longer spoken language contexts has yet to be examined.

4. Conclusions

While not yet a linguistic task, the results of this experiment suggest that listeners' judgments of time intervals are influenced by the magnitude of pitch changes across these intervals. As such, it appears that the auditory kappa effect obtains in speech perception. Further investigation is warranted to determine the extent to which phenomena such as the kappa effect, involving a degree of time-pitch interdependence in perception, affect judgments of linguistic prosody as well.

5. Acknowledgements

This work was supported by NSF grant #1023853.

6. References

[1] Yu, A., "Tonal effects on perceived vowel duration." In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Eds.), *Papers in Laboratory Phonology* (Vol. 10). Berlin: Mouton de Gruyter, 2010.

[2] Cumming, R. "The effect of dynamic fundamental frequency on the perception of duration." *JPhon*, 39(3): 375–387, 2011.

[3] Henry, M. *A Test of an Auditory Motion Hypothesis for Continous and Discrete Sounds Moving in Pitch Space*. Dissertation. Bowling Green State University, 2011.

[4] Cohen, J., Hansel, C., & Sylvester, J. "A new phenomenon in time judgment", *Nature*, *172*: p. 901, 1953.

[5] Sarrazin, J., Giraudo, M. & Pittenger, J. "Tau and Kappa effects in physical space: the case of audition." *Psychological Research*, 71(2): 201–218, 2005.

[6] Grondin, S., Kuroda, T. & Mitsudo, T. "Spatial effects on tactile duration categorization." *Canadian Journal of Experimental Psychology*, 65(3): 163–167, 2011.

[7] Cohen, J., Hansel, C. & Sylvester, J. Interdependence of temporal and auditory judgments. *Nature*, *174*: 642–644, 1954.

[8] Shigeno, S. "The auditory tau and kappa effects for speech and nonspeech stimuli", *Perception & Psychophysics*, 40(1): 9–19, 1986.

[9] Shigeno, S. "The interdependence of pitch and temporal judgments by absolute pitch possessors." *Perception & Psychophysics*, 54(5): 682–692, 1993.

[10] Crowder, R. & Neath, I. "The influence of pitch on time perception in short melodies", *Music Perception*, 12(4): 379–386, 1995.

[11] MacKenzie, N. *The kappa effect in pitch/time context*. Dissertation. Ohio State University, 2007.

[12] Henry, M. & McAuley, J. "Evaluation of an imputed pitch velocity model of the auditory kappa effect", *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2): 551–564, 2009.

[13] Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. "Segmental durations in the vicinity of prosodic phrase boundaries", *JASA*, 91: 1707–1717, 1992.

[14] Bates, D. & Maechler, M. "lme4: Linear mixed-effects models using S4 classes." R package version 0.999375-32, 2009.

[15] Baayen, R., Davidson, D. & Bates, D. "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language*, 59: 390-412, 2008

[16] Beckman, M., & Ayers Elam, G. *Guidelines for ToBI Labelling* (version 3, March 1997).

[17] Ladd, D. "Declination 'reset' and the hierarchical organization of utterances." *JASA*, 84: 530-544, 1988.

[18] Féry, C. & Truckenbrodt, H. "Sisterhood and tonal scaling", *Studia Linguistica*, 59(3): 223-243, 2005.