TIMING OF INTONATIONAL EVENTS IN AUSTRALIAN ENGLISH

Janet Fletcher[a] and Jonathan Harrington[b]

a)Department of Linguistics and Applied Linguistics
University of Melbourne

b)Speech, Hearing and Language Research Centre
Macquarie University

ABSTRACT- There are two competing views of pitch accent timing in English. One suggests pitch accents are timed as a proportion of syllable rhyme duration. Another suggests pitch accent timing is better modelled as an absolute time delay from the onset of syllables. Two corpora were analysed to test which model best fits the timing of prenuclear and nuclear pitch accents in Australian English. Results suggest that syllable onset as well as rhyme duration may play an important role in determining pitch accent timing.

INTRODUCTION

It is well known that prosody and its major constructs for English - stress and phrasing - may be one aspect of spoken language production that facilitates comprehension of spoken language by providing important cognitive processing landmarks to help a listener detect key words or crucial grammatical junctures (Beckman, 1996). Furthermore, these processing land marks can be clearly discernible from the acoustic signal. For example, linguistically focused words often contain prominent intonational peaks and exhibit significant acoustic lengthening ; and phrase edges are often marked by pronounced lengthening (eg. Fletcher and McVeigh, 1993).

A number of studies have found that crucial points of an F0 contour such as the point of maximum F0 in a nuclear accented syllable or point of pitch movement change is positively correlated with the length of syllables or syllable rhymes (Silverman and Pierrehumbert, 1990). Fletcher and Harrington (1993) also note that intrinsic vowel length can interact with pitch accent timing. They note peak retraction (ie. leftward shift) in nuclear accented syllables with rhymes consisting of short versus long vowels. Furthermore they note peak retraction at fast versus slow tempi. Silverman and Pierrehumbert further suggest that F0 peaks are aligned phonetically with reference to the "sonority" profile of a syllable to augment the perceived prominence of a rhythmically strong syllable. By sonority profile they refer to the opening and closing gestures for the syllable that give rise to an increase and decrease in sonority. The F0 gesture is aligned to the entire profile and not just aligned to vowel onset. They suggest that pitch accent timing is therefore relative to the overall duration of the syllable rhyme and should not be modelled as a fixed interval from the onset. In a related vain, they claim that the peak delay in prenuclear accented syllables is greater than in nuclear accented syllables, presumably to "make room" for the realisation of final falling tunes that mark the phrase edge.

In a further study of pitch accent timing in English, van Santen and Hirschberg, (1994) suggest that peak delay in nuclear accented syllables is also affected by length of syllable onsets. Peak delay is greater when syllable onsets are long. Conversely, Rietveld and Gussenhoven (1995) found that sonorant consonant clusters in the syllable onset shifts targets to the left of the accented syllable. Even in languages not normally considered to be stress accented, similar timing effects to those reported by van Santen and Hirschberg have been noted. For example, in a study of Mexican Spanish, Prieto, van Santen and Hirschberg (1995) found that as vowel and onset durations increased in accented syllables, so too did peak delay. However, proximity to a prosodic boundary (word or phrasal) reversed the effect, echoing Silverman and Pierrehumbert's findings for American English.

In many speech synthesis systems pitch accents are timed as a fixed percentage of the total vowel. Alignment of the F0 targets in pitch accented syllable is somewhat insensitive to the segmental make up of syllables and in particular syllable rhymes. Furthermore, Rietveld and Gussenhoven suggest that tune identity is affected when syllables composed of different segments are synthesised with a fixed peak delay from the onset of a vowel. On the basis of the above results, it seems apparent that segmental factors need to be considered in phonetic implementation rules for intonation components of text-to-speech synthesis systems in stress-accent languages.

This paper compares timing of nuclear and pre-nuclear accents in a corpora of spoken Australian English. We set out to investigate how intrasyllabic durations influence the timing of pitch accents, and how proximity to prosodic boundaries can influence pitch accent placement.

METHOD

Speech data from the Australian National Database of Spoken Language (Millar et al. 1990) formed the corpora for this study. The first corpus consists of 680 phonetically dense and balanced sentences recorded for one male speaker of Australian English. The second corpus consisted of one map-task dialogue (McAllister et al., 1990) also recorded as part of the ANDOSL initiative. The sentences and dialogue were recorded and digitised at 20 Khz using ESPS/ Waves + running on Sun Work station at the Speech Hearing and Language Research Centre, Macquarie University. The sentence data were segmented and annotated (see Croot et al. 1992), following standard acoustic phonetic segmentation procedures. An intonational analysis of the corpora was then performed. The theoretical framework adopted in this study is the modified Pierrehumbert intonational model , called ToBI (tones and breaks indices) described in Pitrelli et al., (1993). This model differs from "dynamic" or contour-based intonational descriptions of English proposed by the British School (eg. O'Connor and Arnold, 1971) in that the intonation contour of an utterance is broken down into a sequence of tones, H (high) or L (low) aligned to a rhythmically prominent syllable in a prosodic word (Pitch accents), or to the edge of a larger phonological grouping, the intonational phrase (Boundary Tones). The tone targets occur within a given pitch range that varies from speaker to speaker, and can be modified depending on a combination of phrase-internal pitch range modifications (upstep, downstep, final lowering) and discourse-related factors (eg. position of the phrase in the overall discourse - the so-called paragraph effect, "speaking up") or socio-phonetic factors.

Three tonal events were labelled - pitch accents, phrase accents and boundary tones following ToBI conventions. The Mu+ database management system (Harrington et al., 1993) was used to retrieve and analyse all instances of H*, L+H*, !H* and L+!H* pitch accents and their associated syllables (nuclear and prenuclear) in the first corpus. There were over instances 2761 simple and bitonal H* pitch accents. The former were grouped according to whether they were nuclear or prenuclear,and according to length and segment makeup of onsets and codas. In the second corpus all instances of simple H* accents were retrieved. As the second corpus was labelled at the word, tone, and break index level, it was not possible to analyse pitch accented syllables according to their segmental make-up.

The following measurements were derived from the first corpus using mu+: overall syllable duration and duration of syllable onsets (voiced sonorants vs voiceless clusters), and peak delay (ms) of the F0 target from the onset of the accented syllable. An ANOVA was performed on the data to test the significance of any differences in peak delay and Pearson Product Moment Coefficients were performed on the data to test the degree of correlation between peak delay and overall syllable duration. In the second corpus, we compared peak delay in nuclear versus prenuclear syllables.

RESULTS

Figure 1 plots peak delay as a function of syllable duration across corpus 1. There was a significant correlation between overall syllable duration and pitch accent timing for all accent types($t=71.013$, $df=2761$, $P<0.0001$; $r=0.804$). Breaking the data down into accent type, simple versus bitonal accents showed positive correlations ($t=28.74$, $df=1728$, $p<0.0001$; $r=.568$ for H* accents: $t=14.31$, $df=801$, $p<0.0001$; $r=.451$ for L+H* accents). There were no significant differences in peak delay among bitonal and simple accents.

Within the category of medial prenuclear H* accented syllables, there was a strong correlation between peak delay and the location of the F0 peak ($t=20.95$, $p<0.0001$; $r=.769$). However, as shown in Figure 2, F0 peaks in prenuclear H* accented syllables occurred only slightly later than in nuclear accented syllables that were also phrase final (190 ms vs 185 ms). Peak delay accounted for .7 of overall syllable duration in prenuclear syllables and .4 of overall syllable duration in phrase final nuclear syllables.

Comparing phrase-final nuclear accented syllables with complex and simple onsets (Figure 2) , peak delay increased with complexity of syllable onset ($F=8.488$; $p<0.005$). Peak delay was on average 150ms in accented syllables (.42 of overall syllable duration) with simple onsets compared to 192 ms in syllables with complex onsets (.526 of overall syllable duration). Syllables

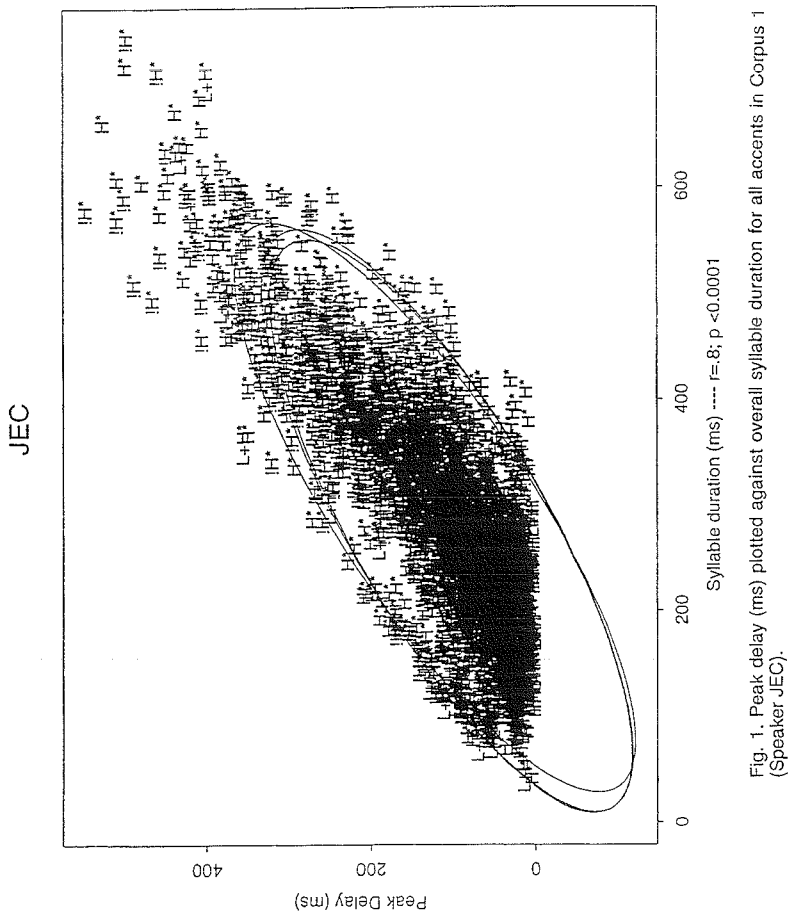

Fig. 1. Peak delay (ms) plotted against overall syllable duration for all accents in Corpus 1 (Speaker JEC).

with simple onsets also showed no significant correlation between duration and peak delay ($p>0.05$). Conversely, in syllables with complex onsets, overall syllable duration correlated positively with peak delay ($t=2.28$, $p<0.05$; $r=.318$).

Figure 3 compares peak delay in Nuclear phrase-final syllables and Prenuclear medial syllables in one map dialogue (the "Leader" role") No significant peak retraction was observed in these data ($p>0.05$). Differences were only of the order of 8ms on average.

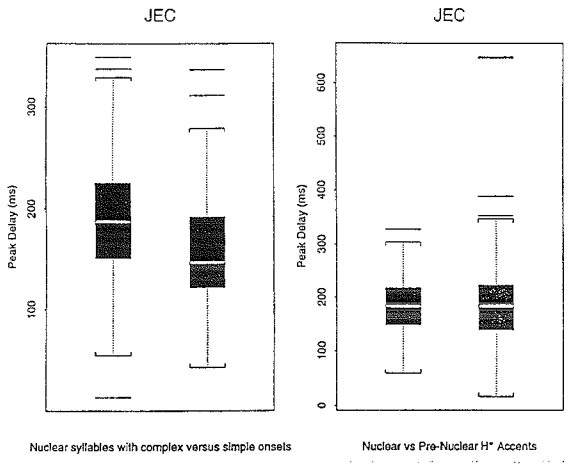

Fig. 2. Comparison of mean peak delay (ms) in Nuclear syllables with complex vs. simple onsets, and in Nuclear vs Prenuclear syllables in Corpus 1 (Speaker JEC)

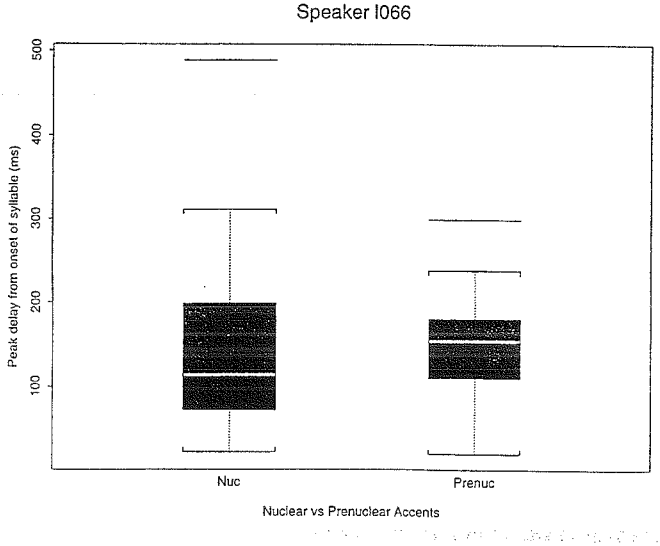

Fig. 3. Mean peak delay in Nuclear vs. Prenuclear syllables in Corpus 2 (Speaker I066).

DISCUSSION

On the whole, peak delay is correlated with syllable duration in nuclear and prenuclear syllables. As syllables increase in duration, peak delay also increases. Peaks are not located at some fixed distance from syllable onset, particularly when the accents are not nuclear or phrase final. Unlike in Silverman and Pierrehumbert, analysis of the first corpora in this study showed that peak position is

not significantly retracted when the accent is adjacent to a phrase boundary unless syllables consist of simple onsets. Phrase-final nuclear accents undergo leftward pull in syllables with shorter as opposed to longer onsets. In our second corpus we were not able to control for segmental composition of accented syllables. This might account for the lack of significant peak retraction.

These findings concur with those reported by van Santen and Hirschberg (1994), and Prieto et al., (1995), but differ from those reported by Rietveld and Gussenhoven (1995) for Dutch. The latter found that complex onsets (containing a sonorant segment) should pull a peak leftwards rather than rightwards. Nonetheless, the results of this study and those reported above for American English, do suggest that alignment algorithms for tonal events (and in particular, pitch accents) in speech synthesis systems should be sensitive to intrasyllabic timing factors.

NOTES

This research was supported by two ARC small grants to the first author . The assistance of Dailan Evans and Chie Hama in labelling the second corpus is gratefully acknowledged.

REFERENCES

Beckman, M. (1986). Stress and non-stress accent. Dordrecht:Foris

Croot, K. , Fletcher, J., and Harrington, J. (1992).Levels of segmentation and labelling in the Australian National Database of Spoken Language. In Eds.J.Pittam and J.Ingram. *Proceedings of the Fourth Australian International Conference on Speech Science and Speech Technology*, 86-91.

Fletcher, J. And McVeigh A. (1993). Segment and syllable duration in Australian English. *Speech Communication*. **13**; 355-365. Amsterdam:Elsevier

Fletcher, J. and Harrington, J. (1993). Pitch accent timing and jaw lowering in Australian English. Paper delivered at the 3rd Tobl workshop, Columbus, Ohio.

Harrington, J., Cassidy, S. , Fletcher, J. and McVeigh, A. - The mu+ system of database analysis. *Computer, Speech, and Language*, **7**, 305-331. London:Academic Press (1993)

McAllister, J.M., Sotillo, C., Bard, E. and Anderson, A. (1990). Using the map task to investigate variability in speech. *Occasional Paper. Department of Lingusitics, University of Edinburgh.*

Millar, J., Vonwiller, J. Harrington, J. and Der mody, P. (1994). The Australian National Database of Spoken Language, *Proc. ICASSP-94,* 197-100.

O'Connor, J.D. and Arnold, G. (1971). Intonation of Colloquial English. London: Arnolds

Pitrelli, M. Beckman, M. and Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the 1994 International Cobference on Spoken Language Processing.* 123-126.

Prieto, P., van Santen, J. and Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429-451.

Rietveld, T. and Gussenhoven, C. (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics*, 23, 375-385.

Silverman, K. and Pierrehumbert, J. (1990): The timing of prenuclear high accents in English. In. J. Kingston and M.Beckman(eds.) *Papers in Laboratory Phonology I: between the grammar and physics of speech.* CUP: Cambridge pp72-106.

van Santen, J. and Hirschberg, J. (1994). Segmental effects on timing and height of pitch contours. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Vol. 1, 719-722.