Perceptual Centres in Speech - An Acoustic Analysis.

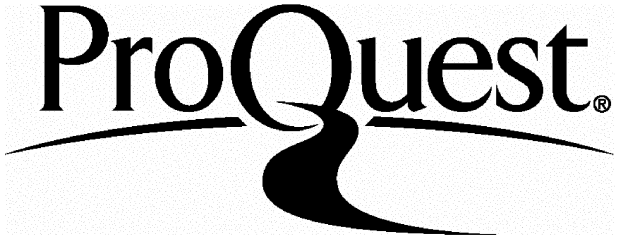
by

Sophie Kerttu Scott

Thesis Submitted for the Degree of Doctor of Philosophy

University College London

September 1993



Abstract

Perceptual centres, or P-centres, represent the perceptual moments of occurrence of acoustic signals - the 'beat' of a sound. P-centres underlie the perception and production of rhythm in perceptually regular speech sequences. P-centres have been modelled both in speech and non speech (music) domains. The three aims of this thesis were to a) test out current P-centre models to determine which best accounted for the experimental data b) to identify a candidate parameter to map P-centres onto (a **local** approach) as opposed to the previous **global** models which rely upon the whole signal to determine the P-centre, c) the final aim was to develop a model of P-centre location which could be applied to speech and non speech signals. The first aim was investigated by a series of experiments in which a) speech from different speakers was investigated to determine whether different models could account for variation between speakers b) whether rendering the amplitude time plot of a speech signal affects the P-centre of the signal c) whether increasing the amplitude at the offset of a speech signal alters P-centres in the production and perception of speech. The second aim was carried out by a) manipulating the rise time of different speech signals to determine whether the P-centre was affected, and whether the type of speech sound ramped affected the P-centre shift b) manipulating the rise time and decay time of a synthetic vowel to determine whether the onset alteration was had more affect on P-centre than the offset manipulation c) and whether the duration of a vowel affected the P-centre, if other attributes (amplitude, spectral contents) were held constant. The third aim - modelling P-centres - was based on these results. The Frequency dependent Amplitude Increase Model of P-centre location (FAIM) was developed using a modelling protocol, the APU GammaTone Filterbank and the speech from different speakers. The P-centres of the stimuli corpus were highly predicted by attributes of the increase in amplitude within one output channel of the filterbank. When this was used to make predictions of the P-centres for all the stimuli used in the thesis, 85% of the observed variance was accounted for. The FAIM approach combines aspects of previous speech and non speech models (Gordon 1987, Marcus 1981, Vos and Rasch 1981). P-centre were thus modelled in a non speech specific, local manner.

Acknowledgements

Thanks to: the Speech Group at UCL for providing the all help I needed - Pippa Bark, Peter Howell my supervisor, Stevie Sackin (for his support and continual assistance) and Keith Young; Richard Baker at UCL Phonetics Dept for filter advice; David Galbraith for getting me here in the first place; to all the subjects who took part, especially David Clynch who was a star; to Gerry Lame and John Morton for very helpful comments; to Paul Burgess and A.R. Jonckheere for statistical advice; to Precilla Choi, Tom 1 Hartley, Linzi Edwards and William Curran for being excellent office-mates; to Suzanne McKeown, Tamar Pincus, Simon Richardson, John Draper, George Houghton and all the other postgrads and panto-queens for making this a great three years; to Colin and Christine Scott for their encouragement; and all my love and thanks to Tom Manly whose support, comments, counsel and care made it possible.

This thesis is dedicated to Winifred Phillips and Dorothy Scott.

Contents

Page

Chapter One: *Perceptual centres - the psychological moments of occurrence of acoustic signals* 6

Chapter Two: *Perceptual consequences of the intensity and time variation in signals* 38

Chapter Three: *Outline of thesis* 57

Chapter Four: *Methodology - how to determine P-centres in perceptual tasks* 65

Chapter Five: *Experiment Two - Perception and production of rhythmic sequences* 96

Chapter Six: *Experiment Three - Different speakers, different P-centres?*116

Chapter Seven: *Experiment Four - Does infinite peak clipping affect the P-centres of speech signals?* 147

Chapter Eight: *Experiment Five - Amplitude envelope and P-centre location in production and perception* 172

Chapter Nine: *Experiment Six - Effect of stimulus rise time on P-centre location* 192

Chapter Ten: *Experiment Seven - A comparison of the effect of ramping the stimulus onset or offset on P-centre; are onset events more important?*....225

Chapter Eleven: *Experiment Eight - Effect of vowel duration on P-centre location* 249

Chapter Twelve: *Local model of P-centre location* 271

Chapter Thirteen: *Conclusions* 297

Page

References 314

Appendices:

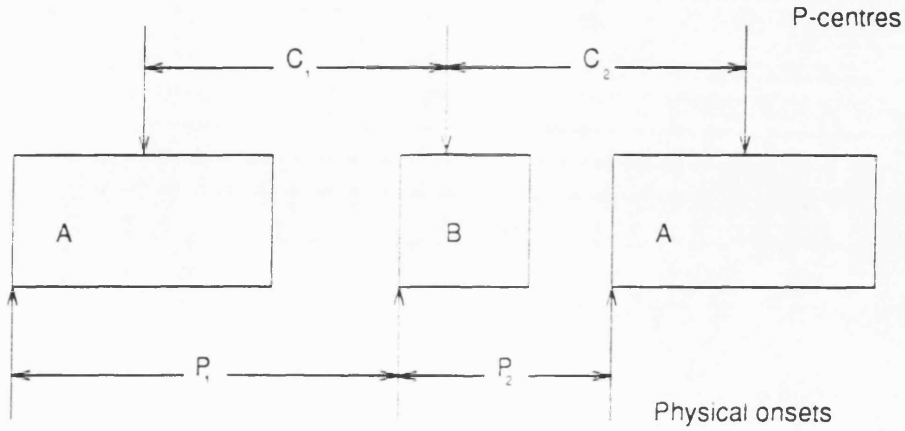
Marcus Model implementation 321
Howell Model implementation 324
Vos and Rasch Model implementation 325
P-centre algorithm implementation 327
Dynamic rhythm setting task code 337
Oscillogram and spectrogram of "wa" stimuli 344
Oscillogram of reference stimulus 345
List of Tables 346
List of Figures 348

Chapter One
Perceptual centres - the psychological moments of occurrence of acoustic signals.

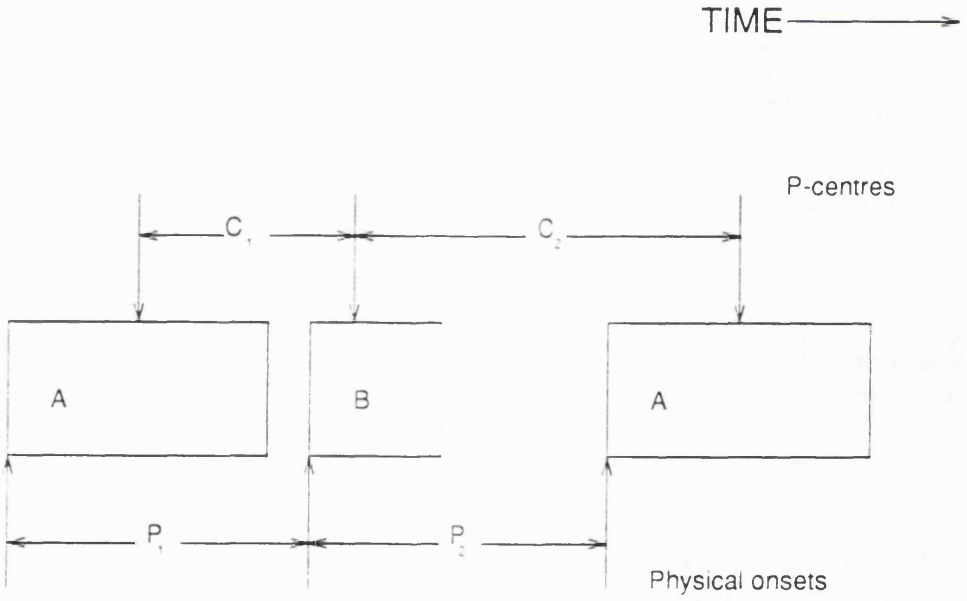
Abstract
Perceptual centres, or P-centres, correspond to the psychological moment of occurrence of acoustic signals, and have been hypothesized to form rhythmic centres in perceptually regular rhythms. In this section there will be a description of the perceptual centre phenomena, and the different theoretical approaches to the topic. The relevant experimental work will be addressed in detail, as will the strengths and weaknesses of the various P-centre models.

1.1 Introduction

In a "Theoretical note" Morton, Marcus and Frankish (1976) presented a short treatise designed to "introduce a new term into the language of....speech perception" (p405). They wished to describe and define a phenomenon which they considered to be "obvious" (ibid). This is that an acoustic signal is not undifferentiated over time, but has a beat, or centre. They stated that the perceptual centre - or P-centre - of a signal (such as a monosyllabic word) "corresponds to its psychological moment of occurrence" (ibid). This phenomenon became noticeable when they were originally preparing speech stimuli to be presented to a subject in a memory experiment. They wanted to control for the rate of presentation, and thus spaced the acoustic items at physically equal onset to onset intervals. They noticed that this physical regularity did not lead to perceptual regularity; in fact the sequences were not perceived as at all even. This led them to question what is regular in an even list, if it is not the signal onsets? They thus defined the P-centre as *what is evenly timed in a perceptually even sequence*. Figure 1.1 shows the difference between perceptual and physical isochrony.



Perceptual isochrony: P-centres are evenly timed ($C_1 = C_2$), physical onsets are unevenly timed ($P_1 > P_2$). The sequence sounds regular.



Physical isochrony: Physical onsets are evenly timed ($P_1 = P_2$), P-centres are unevenly timed ($C_1 < C_2$). The sequence sounds irregular.

Figure 1.1 the difference between perceptual isochrony (top) and physical isochrony (bottom).

In an experiment designed to establish which aspects of a signal subjects line up regularly to produce an even sequence, they allowed subjects to adjust the intervals between two repeating stimuli, until they acheived perceptual isochrony. Subjects could do this, and were consistent across each other, and the intervals they consistently set deviated significantly from physical isochrony. They used the interval settings that the subjects made to calculate the P-centre of the signals using a least sum of squares fit (see Method chapter for details).

They stated as a null hypothesis that P-centres could be quantified in terms of the acoustic signal, and provided some evidence about this. It was all negative however; P-centres did not correspond to the signal onset, the vowel onset or the point of peak amplitude, neither were they the result of a simple energy integration function (Marcus 1975). Finally they presented evidence that this phenomenon was found when subjects produced even sequences of speech, and hypothesized that analogues of P-centres would be found in the field of music perception and production, for example in the synchronous playing of instruments (here they anticipated the work of Vos and Rasch (1981) on musical beat location - see the next chapter).

Work by Rapp (1971) and Allen (1972), in production and perception respectively, were precursors to this concept definition. Rapp instructed subjects to tap along to a sequence of words to determine where the 'syllabic stress beat' locations were. He found an asynchrony between where people tapped and the physical signal.

In Allen's production experiments subjects produced rhythmic speech in time to a metronome. They would begin to utter the syllable before the metronome click, and the longer the prevocalic consonant cluster, the earlier they would commence their articulation.

1.2 Marcus's Model of P-centre location

Marcus (1981) who had been a co-author of the original P-centre paper published a model of P-centre location. His model specifically discounted any one acoustic event as a determinant of P-centre location, and instead was based upon the whole syllable. His model is thus a **global** model of P-centre location; it does not map P-centres onto a single acoustic event. A model which did adopt such an approach would be a **local** model of P-centre location.

He conducted several experiments in which he varied aspects of the speech stimuli and measured the effect on the P-centres of the syllables. To do this he used a general method similar to that described later in the Methodology chapter. He used an indirect measure of the beat location of signals - the dynamic rhythm setting task. The assumption of this method is that when a listener adjusts the intervals in a sequence of sounds to attain perceptual evenness, they will be using the beats or rhythmic centres of the sounds to set the even rhythm. The pattern of physical intervals that are set can be used to calculate where the rhythmic centres are in the signal.

Thus in each dynamic rhythm setting experiment, each individual stimulus is set to a rhythm against every other stimulus in the experiment. For **n** stimuli, there are **n X n** experimental trials. This method enables the intervals that the subjects set to be used to determine relative P-centres for the experimental stimuli (relative to all the other sounds in the experiment). Marcus's model is based upon actual P-centres, rather than mean offsets from isochrony. The intervals between the repeated stimuli were adjusted by the subjects by turning a knob. The initial interval of the two stimuli was randomised for the start of every trial, so subjects could not be making stereotypical responses based on turning the knob to a certain point. Each combination was presented only once.

1.3 Marcus's experiments

His first experiment determined P-centres for spoken digits, "one" to "nine". In addition he showed no significant difference between subjects, confirming that the P-centre phenomenon is one which is not subject to individual variation. Based on this finding he used only himself as a subject in subsequent experiments.

His next experiment showed that decreasing the duration of a prevocalic consonant has a strong effect upon the P-centre of the syllable. He successively deleted segments of the initial /s/ fricative in "seven". This resulted in a continuum from "seven" through "tseven" to "deven" . This effect of frication duration reduction altering the perceived phoneme, and in a categorical manner is a well established finding (Gerstman, 1957). Marcus observed this phonemic perceptual categorization, but did not test for it formally by using the twin tasks of identification and discrimination.

This experiment showed that the longer the initial consonant of a syllable, the later the P-centre. There were no abrupt shifts in P-centre caused by the categorization of the words in the continuum. The P-centres shifted smoothly, although the physical amount of energy removed at each level differed a lot (due to natural ramping of the stimuli).

He next tested the effect that vowel duration had on the P-centre. He did this by lengthening the vowel in CV syllables (bae, dae, gae); the vowels were pitch synchronously extended. He found a similar effect to initial consonant duration; the longer the vowel, the later the P-centre of the syllable. However the effect of vowel duration was much weaker - P-centres shifted by about a third of the change of the vowel duration. He concluded that P-centres were a result of the entire stimulus, rather than one acoustic property of the stimulus.

In his final experiment he extended this investigation of the effect of post-vocalic duration to syllable final consonants. The spoken word "eight" consists of a vowel followed by a "t" burst. This phoneme is an unvoiced plosive which is produced by fully occluding the oral cavity whilst increasing pressure in the mouth. When the pressure is released by lowering the tongue there is a burst of noise as the trapped air rushes out. Acoustically, there is a period of silence when the mouth is obstructed, between the vowel offset and the noise burst. Marcus increased and decreased the duration of the silence before the burst by 30ms in either direction. He also amplified the final "t" burst by 4.5 and 9dB SPL. He thus varied the duration of the stimulus, and the amplitude profile of the stimulus as two independent variables. He noted that the silence duration modifications were scarcely noticeable, whilst amplifying the "t" burst was very salient. The P-centres of the syllables varied with the silence duration manipulations; the shortened gap duration leading to an earlier P-centre and the lengthened gap leading to a later P-centre. Increasing the "t" burst amplitude had no effect on P-centre, however. In his paper he does not give means and SD's for the raw intervals set; certainly for the amplitude increased "t" burst stimuli the P-centres vary quite a lot.

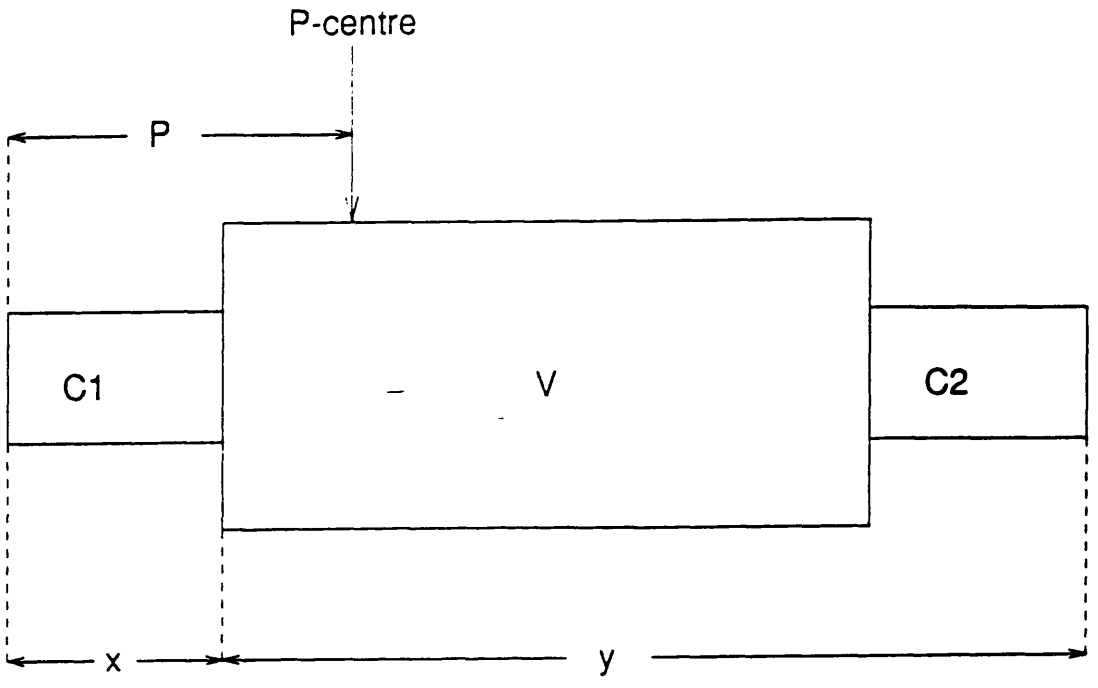
1.4 Marcus's model

Marcus developed a model of P-centre location on these results. It relates P-centre location to the duration of different portions of a syllable, the onset and the rhyme. He defined these acoustically, and weighted the signal onset - acoustic vowel onset portion as more important than the vowel onset - signal offset portion. His model of P-centre location was defined as:

$$\textbf{P-centre} = 0.65(A) + 0.25(B) + K$$

Where

A = duration from signal onset (passes a threshold of -30dB SPL) to vowel onset (defined as the peak increment in mid-band spectral energy 500-1500Hz).



Marcus's model of P-centre location divides the syllable into two portions, here marked x and y. The x represents the prevocalic C1 duration and y represents the V-C2 post-vowel onset to signal offset duration. The Marcus model can be expressed as P = a(x) + b(y) + K, where P is P-centre location relative to stimulus onset , a and b are parameters of the model and K is an arbitary constant.

Figure 1.2. Schematic illustration of the parameters of Marcus's P-centre model (1981) for a C1-V-C2 syllable. (After Marcus, 1981, p252).

B = Duration from vowel onset to signal offset

K = arbitrary constant

Figure 1.2 shows a diagram of the model.

Marcus's model was termed an acoustic model of P-centre location, though it could also be regarded as a phonetic model, since its syllable divisions are phonetically motivated. His model takes the entire signal as having a role in P-centre location, though not all of equal weighting. He explicitly denies thus that any one aspect of the syllable affects P-centre location. He also defined P-centres as free of context sensitivity. His model thus predicts that a syllables P-centre would not vary with the intensity of presentation, not would it vary with the tempo of the sequence not the other sounds in a sequence.

1.5 Howell's Model of P-centre location

Another model of P-centre location was developed by Howell (1984, 1988a). Howell stated that all simple acoustic factors should first be considered as determinants of P-centre location before adopting a complex articulatory approach. Howell defined three criteria for a determinant of P-centre location:

> 1) "it should vary in alignment across stimuli in the same way as the perceptual judgments do" (p429)
> 2) "It should vary in location relative to stimulus onset in the same way that perceptual alignments vary when the acoustic properties of test stimuli are altered "(ibid)
> 3) "The factor should account for ..(P-centres).. in both production and perception" (ibid).

Howell approached the acoustic modelling of P-centres from the finding that the principle acoustic factor associated with the syllable is the amplitude envelope

(Mermelstein, 1975). Previous research into acoustic determinants had looked for acoustic determinants within the syllable itself (eg. vowel onset as reference point). Howell used the amplitude envelope as an acoustic factor associated with the whole syllable.

He highlighted the fact that all of Marcus's experimental manipulations affected the amplitude envelope of the signals concerned, and thus it could be that P-centres are a result of the amplitude envelope contours of a signal. Simply, this would mean that increasing the amount of energy at the onset of a signal would shift the P-centre forward in time; increasing the energy at the offset would shift the P-centre back.

1.6 Howell's experiment

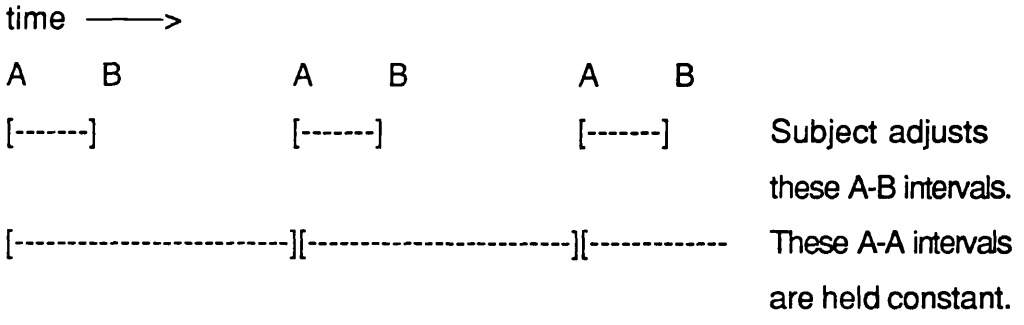
Howell presented evidence to support his model. He created two fricative-vowel stimuli ("sha" and "cha") by ramping the onsets of the syllables while keeping the durations the same. Thus the only difference between the "sha" and the "cha" stimuli was the rise time of the frication (120ms and 40ms respectively). The vowel in each stimulus was the same, in duration and linear tapering.

Howell used these stimuli in a rhythm setting task to see if a different pattern of physical intervals arose from having the two sounds the same ("sha"-"sha") or different ("sha"-"cha"). His method differed from that of Marcus by not having a constant overall tempo; instead the interval between pairs of sounds was altered by a subject until perceptual isochrony was reached. The difference in the two methods is shown schematically in Figure 1.3.

Thus it can be seen that both methods enable subjects to set an even perceptual rhythm with the two stimuli. In the second method the overall tempo of the sequences can change, whereas in the first it cannot due to a constant

A - A interval. P-centres for the signals could not therefore be calculated from these results, but stimuli with different P-centres will still lead to different settings. The fixed interval was 750ms, and the other interval was always determined randomly, to eliminate response bias. Six subjects completed the settings.

Marcus's paradigm:



Howell's paradigm:

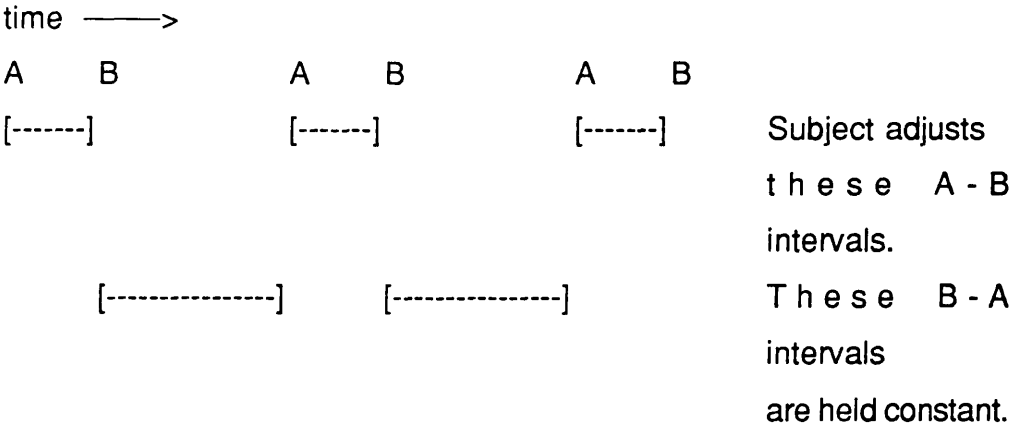


Figure 1.3 Comparison of the intervals aligned in different dynamic rhythm task paradigms (Howell 1984, Marcus 1981)

Howell found there was a significant difference between the intervals when a "sha" was set against a "sha", and when it was set against a "cha" sound. This he attributed to a difference in the P-centres of these stimuli, and since they

differed only in their rise times, he took this as evidence that P-centres are determined by the amplitude envelopes of the signals.

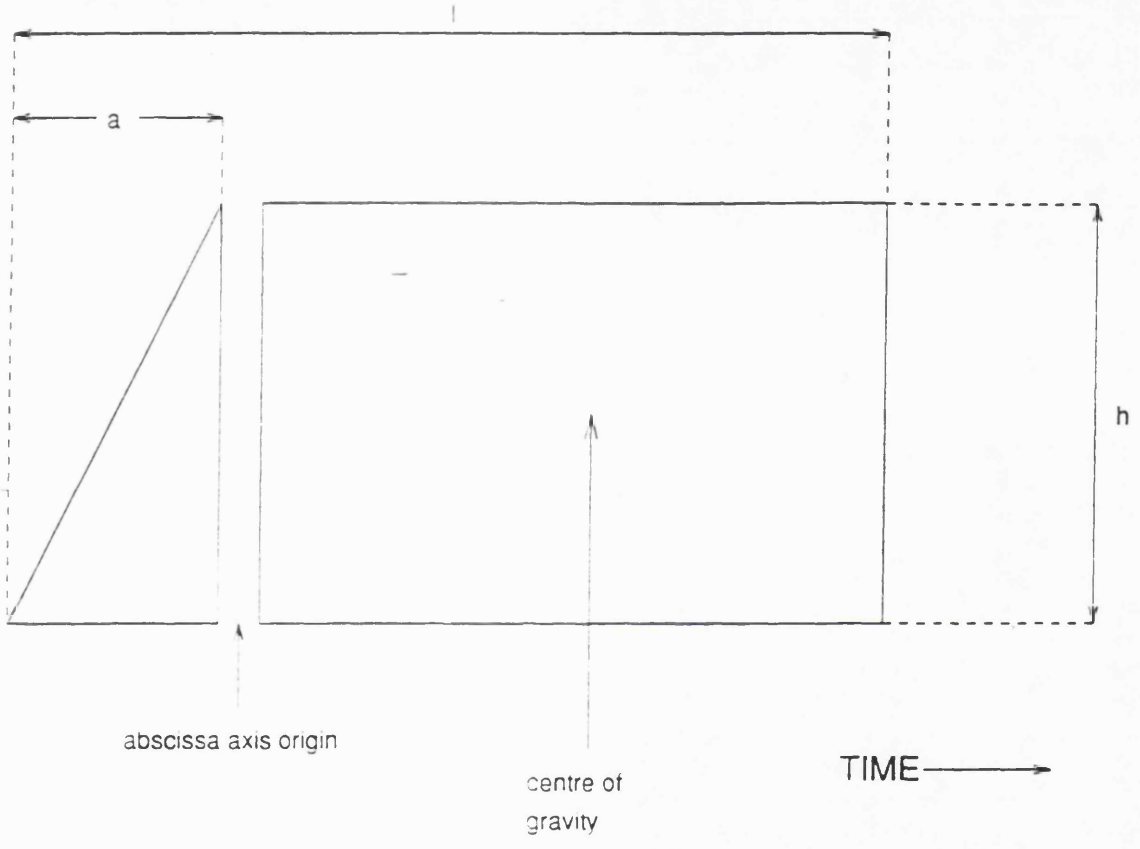
He repeated this experiment using non speech sounds with similar structures. They began with a period of white noise followed by a section of saw-toothed wave excitation. Six subjects (different from the first experiment) performed the experiment). He found the same pattern of intervals resulting in perceptual isochrony. He concluded that P-centres were phenomena not confined to speech sounds.

He later formalized his account of P-centres into a syllabic centre of gravity model (1988a&b). This relates P-centres to the intensity distribution of a signal over its entire duration. Alterations to this distribution will shift the P-centre along with the centre of gravity of the intensity distribution. In terms of the amplitude envelope this is demonstrated in the diagram Figure 1.4.

Thus by the criteria quoted above, the amplitude envelope of a stimulus is an acoustic determinant of P-centre location because

1) the changes he caused to it (ramping the onset) led to an analogous shift in the intervals set by the subjects
2) the changes in intervals set shift in the same direction as the centre of energy (gravity) of the stimulus

The final criterion he set was that the factor must explain the rhythm setting in production as well as perception. In the original 1984 paper he also described data which indicated that speakers produce vowels of different durations at different intervals. This indicates that the duration of a vowel affects its P-centre in production - and the duration is a quality of the amplitude envelope. This goes some way to fulfilling the final criterion:



Schematic representation of amplitude envelope from which a syllabic centre of gravity may be derived. This can be done by integrating the represented energy, or by using the following formula:

Centre of Gravity = ((-0.5ah)(a/3) + ((l-a)/2)(l-a)h)/(0.5ah + (l-a)h).

After the centre of gravity has been obtained, it can be represented on the amplitude envelope by measuring from the onset of the envelope. Thus the syllabic centre of gravity of a signal will shift relative to the physical onset, as a function of the energy distribution of the syllable over time.

Figure 1. Illustration of the syllabic centre of gravity model of P-centre location (Howell 1988a)

3) the factor affects P-centre in production and perception. Speakers, he said, regulate their utterances according to knowledge of the energy distribution of the sounds they are producing.

Howell's model is thus an acoustic account, which involves the amplitude of a signal over its entire duration. Like Marcus's model, it does not appoint one acoustic event as the sole determinant P-centres. It is thus a global model of P-centre location.

1.7 Articulatory Model of Location

While various researchers were modelling P-centres by considering the perceived and produced signals, a different account of P-centre location - different both in the level of analysis and the conceptual framework, was simultaneously being devised in the USA. Carol Fowler and her associates at Haskins laboratories defined P-centres as resulting from articulatory gestures in speech. In this account P-centres are speech specific phenomena.

1.8 Fowler's experiments

Fowler (1979) reported a set of experiments, whose aim was to discount acoustic interpretations of P-centres and establish them as a result of directly perceived articulatory gestures. In her first experiment she reported the inter-onset intervals produced by a speaker who either produced homogenous speech sequences (same word repeated) or sequences of two alternating words, at an even rhythm. Intervals were measured by hand, with a reported sensitivity of +/- 7.5ms. No note is made of how the rhythmicity of the sequences was validated. She found that the intervals between alternate words were significantly different, as established by ANOVA, despite the standard deviations being much larger in the alternating condition than the homogenous

condition (prompting the question over subjective assessment of the evenness). She also found that deviations from absolute isochrony were greater for syllables with longer prevocalic consonant clusters.

The next experiment in the paper investigated the perception of these same sequences by listeners, when presented at the 'naturally' timed intervals, or when adjusted by the addition of silence to be physically isochronous. Listeners chose the natural sequence as more rhythmic at a level above chance. This was obviously not a P-centre determining exercise, but did establish that physically isochronous sequences were not perceived as regular by subjects.

Finally she produced results which indicate that when subjects read out a sentence frame "Jack likes black ----", with one of a number of different monosyllables inserted, the duration of the offset of voicing of the "k" and the physical onset of the next word correlated with the prevocalic duration of the test word. That is, the longer the prevocalic portion, the earlier subjects would start to say it. This accords with the observations about production of speech in metronome speech conditions, and is an indication that P-centres may underlie timing in speech. The experiment was carried out with only one speaker, and presumably the same sensitivity of measurement.

Fowler gave all of these results an articulatory explanation. Speakers performing a 'stress timed speech task' that is, producing even speech, produce articulatory regular speech which results in physical anisochronies but is perceived as regular. A listeners judgment of the evenness of the speech is based upon information about the articulatory timing; the subjects judge the articulatory syllable onsets, not the physical onsets.

Fowler concluded that none of the observed anisochronies are a result of strategies in prosody but are a result of articulatory isochronies which occur

whenever speech is produced. She tested this by testing subjects' vocal reaction times to a list of 20 CV syllables.

The results of this showed an effect of manner of articulation on reaction time, implying that the longer the prevocalic segment, the longer the reaction time. An ANOVA showed a significant effect of manner on reaction time (RT). Scheffe's test indicated that only the difference between the RT's for the affricate and the other RT's was significant. Fowler stated that the rank ordering of the means confirmed her hypothesis. Based on these findings she stated that the reaction times were affected by differences in the manner, class or articulatory characteristics of a consonant that affect the time at which the segment has some acoustic consequence other than silence. Variability due to these factors was responsible for 74% of the variance obtained in the first experiment (when one subject produced sequences of speech). Despite their lack of statistical significance, she states that these factors are "what gives the P-centre phenomenon its character" (1979, p383).

The final experiment, designed to lock P-centres to the onset of articulatory syllables rather than some acoustic attribute, compared the timings of voiced and voiceless pairs of syllables. These pairs are dissimilar acoustically, but similar articulatorily, she stated. Her interpretation of the acoustic models of P-centre location is that the alternation of voiced/voiceless pairs will lead to physical anisochronies; whereas the articulatory account predicts no such anisochronies. This was tested in an experiment in which two subjects read out lists of either alternating or homogenous CV sequences. The alternating sequences contained voiced/voiceless plosive pairs ("ba"-"pa", "da"-"ta", "ga"-"ka"). She found no difference between the intervals when they were measured from burst to burst, but a difference when they were measured from physical onset to physical onset. She found that these results confirmed the articulatory account (although via a failure to reject a null hypothesis).

The articulatory model of P-centres that is stated in this paper relates P-centres to articulatory syllable onset - specifically the articulatory activity associated with the vowel onset. Any acoustic marker, therefore, need not be a specific event, but "a (very large) class of acoustic signals .....that signify the...onset for articulatory activity for a vowel"(p 386). This onset of activity she links to anticipatory coarticulation. She hypothesizes that the "stabilization of a vowel's effect on the acoustic signal (of) the consonant"(p387) marks the P-centre for a listener. Her model of P-centre location is thus, at this stage, a local one, mapping P-centres onto vowel gestures.

1.9 Tuller and Fowler's experiments

The following year Tuller and Fowler (1980) presented evidence of articulatory isochrony in perceptually even but physically isochronous sequences. Their stated aim was to "discover whether any part of the production of psychologically isochronous utterances is isochronous even when the acoustic product departs from isochrony" (p278).

They took EMG recordings of subjects producing rhythmic sequences of homogenous or alternating CV syllables. The syllables were chosen to involve the orbicularis oris muscle for articulation of either the initial consonant or the vowel. The orbicularis oris muscle acts to close or protrude the lips and is associated with the production of labial consonants and lip rounded vowels. The four subject's utterances and EMG measures were recorded on separate channels. No record is given of how the evenness of the produced sequences was validated.

They found that in the alternating conditions, there was a significant deviation from physical isochrony in the acoustic intervals. However for the EMG data they found no significant deviation from isochrony in any condition. The EMG

interval values for each sequence were based upon the average for each sequence - that is, not on the individual EMG trace intervals. Howell (personal communication) has pointed out that this is more likely to result in a lack of anisochrony for the EMG data compared to the raw intervals processed for the acoustic data. Certainly some of the interval difference scores for the EMG data look quite large. They took this finding of a suggestion of muscle activity isochrony as evidence for an articulatory basis for P-centres. They fitted this finding into the motor theory of speech perception (Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967).

Tuller and Fowler (1981) tested the hypothesis that P-centres are articulatory gestures, directly perceived by the listener, in a further experiment. This was designed to show that removing all intensity/time variation in a speech signal does not affect its P-centre. This was congruent with a direct model of P-centre location, since it would predict that such a manipulation would not affect the underlying articulatory gestures and would therefore not affect the P-centre of a syllable.

They tested this by rendering the amplitude envelope of a signal entirely invariant via a process called infinite peak clipping. This means that all the positive amplitude values of a signal are increased to a maximum, all the negative to a minimum. The rise and fall time of the resulting signal are immediate, and the amplitude envelope is rectangular. The duration of the signal and its frequency content are unaffected. The speech remains comprehensible, although it sounds harsh and noisy (cf. guitar fuzz box). Tuller and Fowler recorded perceptually isochronous alternating sequences of syllables from a speaker. This was then infinitely peak clipped by amplifying the signal up to a minimum to create two sets of stimuli - normal and infinitely peak clipped speech sequences.

A visual examination of their stimuli reveals that the speech was not infinitely peak clipped (Howell 1988a&b). The amplitude envelopes are not rectangular, the onsets and offsets are ramped, and the durations seem to be different. This point will be returned to in more detail in Experiment 4.

In common with their previous perceptual work on the P-centres, they did not perform a rhythm setting task to enable them to determine the P-centres of signals. Instead they presented subjects with naturally timed sequences, or artificially rendered physically isochronous sequences. Subjects rated the naturally timed sequences are more rhythmic, whether or not the sequences had been infinitely peak clipped.

Tuller and Fowler concluded, that since gross distortions to the amplitude envelope do not affect the P-centres of signals, that an articulatory account of P-centre location is preferable to an acoustic one.

1.10 Cooper Whalen and Fowler's experiments

Cooper Whalen and Fowler (1986) working from an articulatory basis extended and amplified Marcus's experiments. Marcus (1981) had found that decreasing the duration of an initial /s/ segment shifted the P-centre of the resulting syllables. He mentioned that, as had been found previously (Gerstman, 1957) the perceptual consequence of this was a continuum between /s/ /ts/ and /d/. He did not test this categorical perception. Cooper et al extended Marcus's experiment to distinguish the degree to which the phonetic identity of a syllable initial consonant affects the P-centre of a syllable. They did this by employing proper categorisation/discrimination tasks, small steps between levels of the continua, and phonotactically legal English sounds. They also aimed to test Howell's (1984) hypothesis that the amplitude envelope of a signal determines its P-centre.

They carried out four experiments with the same general pattern. First a forced choice identification task and an ABX discrimination task were performed. This was followed by an 'alignment' task, "designed to measure the relative P-centre location of the test stimuli" (p189). In this the test sound was set to a rhythm against a reference sound (a "ba" syllable 329ms in duration). The A - B interval was adjustable, as in Marcus's paradigm, and the initial A - B interval was always 50ms.

They used two different alignment systems. In one subjects adjusted the A - B interval by pressing keys on a computer keyboard, which altered the interval in steps of 1, 5 or 15ms in either direction. In the second, adjustments were made by turning a knob. The smallest possible adjustment was 12.8ms. They combined the results from the two systems. Both could be criticized however, for being in the first case rather counterintuitive, and in the second not very sensitive.

In the **first experiment** they varied the duration of an initial "sh" consonant to create a continuum which varied across "sha" - "cha" - "ta". They ramped the onset of the frication, but not to a regular amount. The shorter the frication, the steeper the rise time. They found that this manipulation resulted in categorical perception of the "sha" - "cha" - "ta" continuum, and that the offset from acoustic isochrony (which they equate with P-centre) varied linearly with the duration of the frication. The category boundaries caused by the variation in duration did not affect the P-centres abruptly. Therefore, they concluded, the phonetic identity of a syllable does not affect its P-centre location.

In the **second experiment**, they amplified the first result by using a different continuum, one in which the rise time of the stimuli was not affected. Here they created a "sa" - "sta" continuum by inserting 10ms sections of silence between the frication and vowel onset of a natural "sa" syllable. The resulting perception

was categorical (as shown by identification and discrimination tasks). They found the same relationship between the amount of inserted silence and the offset from isochrony as in the first experiment. Both experiments indicate a 1ms shift in P-centre location for a 1ms experimental manipulation. They incorrectly interpreted this as disproving Howell's amplitude envelope model. This duration extension affected the amplitude envelope, and the P-centre, in the precise way Howell's model predicts.

The **third experiment** controlled for the effects of overall syllable duration by offsetting the amount of silence insertion by excising an equal amount of frication. They found that this manipulation did not affect the P-centre of the resultant syllables, although categorical perception was still observed. The linear shift of offset observed in Experiment 3 was not therefore due to the gap duration of the syllable.

In the **fourth, final experiment** they varied gap duration again, but this time offset the duration by excising portions of the vowel. This led to categorial perception of the continuum, and a linear shift in offset with gap duration. The longer the gap duration, the larger the offset from isochrony. The slope of the relationship is significantly more shallow than in Experiments 1 and 2. They interpret this as being in line with Marcus's original model; the increase in prevocalic duration caused by the insertion of silence shifts the P-centre back in time; the shortening of the vowel shifts the P-centre forward in time. Since the prevocalic portion is more heavily weighted than the vowel portion in P-centre determination, the P-centre overall shifts back in time, but not as much as it would had the vowel remained constant. They did not test this by implementing Marcus's model to see what predictions it made.

They summarise the findings of all the experiments thus; the P-centres of stimuli are not affected by the phonetic identity of the prevocalic segment, or

obvious acoustic properties of the signal such as overall duration or gap duration (despite the large effect of gap duration). They also found that these results showed that an amplitude envelope model (Howell, 1984) did not account for shifts in P-centre location, since when the "onset envelope" of a syllable remained unaltered as in Experiments 2 and 4, the P-centres still shifted. This is based on an incorrect interpretation of Howell's model which applies to the whole envelope, not simply the onset.

What existing evidence was published (Howell 1984, Vos and Rasch 1981) showing the effects of stimulus rise-time on P-centres, they stated was only relevant to non-speech stimuli, and therefore not evidence which could account for P-centres in speech sounds. They place P-centres firmly in the arena of phenomenon specific to speech. They do however accept that all of their results can be interpreted in both an articulatory and an acoustic manner.

Cooper Whalen and Fowler (1988) conducted experiments to show that a syllable's rhyme acts as a unit in its effect on P-centre location, with a much smaller magnitude than the effect of the prevocalic segments. The articulatory model of P-centre location had shifted ground to be more of a global model, since instead of testing for the effect of vowel gestures on P-centre location, their investigations were directed towards the effect of segmental durations on P-centre location.

Their **first experiment** tested whether the weak effects of vowel duration were due to it being in the syllable rhyme, or due to it being further away from the syllable onset. This was done with two continua. The first was constructed by excising whole pitch periods from a natural vowel, so that there were seven levels of vowel duration. The second was created by appending natural /s/ frication onto the vowels of different durations created above. There were thus

"sa" and an "a" continua which were perceived by the authors as sounding natural.

These stimuli were used in an alignment task, setting the stimuli to a rhythm against a reference sound (which was a 329ms /ba/ sound). Subjects adjusted the intervals, which were initially always 50ms, using keys which altered the interval by 1, 5 or 15ms in either direction.

The results of this experiment are discussed in more detail in experiment 9. They are very unusual for P-centre experiments in that they show none of the consistency between subjects which is common (and indeed, possibly an assumption about P-centres). The subjects show very different results; the naive subject MRS showing a similar trend of smaller offsets with shorter vowels for both continua; another subject CAF (one of the authors) showed a much steeper relationship between offset and vowel length, and the third subject, AMC (another author) showed diametrically opposite effects of offset according to continuum, and no effect of vowel duration.

In the face of such disparate results, it might be suggested that not much can be concluded about vowel duration. Cooper et al state however that since no subject showed an interaction between continuum type with vowel duration, that the effect of vowel duration is not due to the temporal location within the stimulus.

The **second experiment** was to determine whether the duration of any segment in the syllable rhyme has the same effect as vowel duration. To do this they created two more continua both based on the syllable "at". In the first continuum the vowel duration was decreased by excising whole pitch periods, to seven levels of duration. In the second continuum, the same stimuli were used, but with silence inserted in the period of closure to compensate for the

reduction in the duration of the vowel. In the second continuum therefore, the overall duration of the syllable is constant. If the syllable rhyme behaves as a unit, then there should be a shift in the P-centre of the first continuum with vowel duration, but not the second. A centre of gravity account (1988a) would predict "little difference between the two continua since *the silence and release burst add little to the amplitude profile of the utterance*" (p28) (my italics). Again the results were analyzed separately for each listener, since the degree of variation for each subject was very large.

The naive subject, MRS, showed no effect of continuum type in ANOVA, but a significant effect of vowel duration; from examination of the published figures this effect of vowel duration was not to be in the predicted direction (shorter vowel leading to smaller offset). Certainly the shift in offset is very small for both continua.

The second subject, AMC (an author) who in the first experiment showed no effect of vowel duration in this experiment showed an effect of vowel duration on the offset - shorter vowel, smaller offset. There was no effect of continuum type.

The final subject CAF (an author) showed the clearest results, with an effect of continuum type, vowel duration, and an interaction all at significance. The offsets of both continua get smaller as the vowel duration decreases, the duration constant continuum showing greater offsets and a shallower slope. The result of this subject alone confirm the hypothesis that the syllables rhyme acts as a unit on P-centre location.

The results of the other two subjects, who showed no effect of continuum type, are hard to dismiss. These results could be explained by other P-centre models, for example the centre of gravity account. The sheer variance between

the offsets of the three subjects is a source of problems for the whole experiment. Subject MRS set intervals which varied from isochrony between -15ms to -40ms. Subject AMC varied from -120m to -60ms. Subject CAF from -80ms to +5ms. This is simply too great a range of responses to use to reach any conclusions. They might wish to claim that the subjects are alike enough in their responses that they are setting their intervals to common trends, but that the range of the offset can vary. In this case, they need to establish theoretically, why the trend remains the same while the range varies, especially since similar experiments have shown no such degree of difference between subjects. The alternative hypothesis is that the experimental results are to unreliable to base a finding upon. Responses which vary in one dimension - range - could theoretically vary in another.

Instead of either of these conclusions, Cooper et al state that the contribution of segments in the rhyme to P-centre location varies across listeners, and regard this as the basis for possible future work. Since Marcus carried out his experiments mainly on himself, the effects of vowel/rhyme duration on P-centres across subjects cannot be verified.

1.11 Further work on P-centres

Marcus's experimental findings were replicated by researchers working with Dutch spoken digits (Eling, Marschall and van Galen 1980). Dutch, like English, is a so-called 'stress timed' language. This is a reference to the supposed timing of a language, stress timed languages being produced with 'isochronous' inter-stress intervals. Syllable-timed languages such as French are considered to be timed with 'isochronous' inter-syllabic intervals. There is some evidence that these timing variations are more in the listener's ear than the physical timing of the language (Lehiste 1975). Can the P-centre hypothesis be applied

to languages which vary in the manner in which they are considered to be timed?

Hoequist (1983) investigated the generality of the P-centre phenomena to different languages. In addition to a so-called stress-timed language (English), he analyzed the production of rhythmic speech from Japanese speakers (a mora timed language), and Spanish (a syllable timed language). To reduce second language problems, his subjects spoke nonsense syllables to an even rhythm. He found the expected anisochronies produced by all his speakers, and concluded that "..P-centres (are) an aspect of syllable production, and ... thus expected to be universal" (1983 p375).

Howell (1984) reported that in production, speakers systematically vary their timing of vowels with different durations to acheive perceptual isochrony. This finding was replicated by Fox and Lehiste (1987). They found that as a speaker produces a longer vowel, the 'stress beat' (analogous to the P-centre in definition) of the syllable moves away from the vowel onset. They found that these differences between tense (long) and lax (short) vowels disappears in a perception task, if the durations of the vowels are equalized. They concluded from this result that the stress beat was influenced by the duration of the vowel. Spectral cues were not leading to the syllable being *expected* to be longer, as they had suspected. Thus the subjects, recognising a tense vowel, did not make rhythm setting judgements depending on how long they anticipated the vowel to be. They concluded that the stress beat of a token depends upon its entire structure. Their experiment thus supports a global model of P-centre location.

Evidence to support the articulatory approach to P-centres was provided by Tye-Murray, Zimmermann and Folkins (1987). They showed that the physical anisochronies which lead to perceptual regularity when speakers produce

rhythmic speech are also shown when prelingually deaf adults produce regular speech sequences. They interpreted these results as indicating that the deaf speakers were producing the vowel articulations regularly without acoustic information, and thus P-centres should be considered as articulatory gestures rather than attributes of the signals.

1.12 Another model of P-centre location.

In contrast to all the approaches to P-centres described in this chapter, Pompino-Marschall (1989,1992) developed a model of P-centre location that is "purely psychoacoustic" (1989 p175). His experimental findings concluded that, unlike Marcus's and Fowler's findings, the prevocalic and rhyme sections of a syllable do not act independently on the P-centre. This is in contradiction to Cooper et al (1988) who stated that the syllable's rhyme does affect the P-centre as a unit. Pompino-Marschall provided evidence that the vowel and final consonant duration are different in their effects; also both effects are non linear. His final finding was that phonetically different syllables affect the P-centre differently, and P-centre shifts can be induced without temporal variation in the signals.

His model is a psychoacoustic one which uses thresholds within loudness functions within each critical band; it also uses temporal weighting processes between these 'partial events' and integration processes at different levels. The model is complex, but can be generally expressed as determining a perceptual syllabic onset and a syllabic centre of gravity, from an integration of the partial events.

Pompino-Marschall's model is used to argue against a model of P-centres based on segmental durations within a syllable, and in favour of an account

which is applicable to both speech and non speech sounds. This model uses energy distribution within frequency channels to predict P-centre location. Unlike any other model of P-centre location, this model is structurally dependent upon the identification of two perceptual occurrences within a syllable, the perceptual onset and the perceptual centre of gravity, which are in turn influenced by the entire syllable. The Pompino-Marschall model is thus 'semi-local'; it maps P-centres onto two discrete events, but events that are not independent of the whole signal in terms of their calculation.

The knowledge that acoustic signals have a perceptual onset which is not the same as their physical onset is as old as the psychophysical tests designed to examine the relationship between the two domains. More recently it has been pointed out that a signals perceptual onset is not necessarily the same as its perceptual moment of occurrence, or beat (Gordon 1987). Whilst this separation of physical and perceptual events is clear, it does not necessarily follow that the P-centre must depend upon the perceptual onset.

Pompino-Marschall (1989) states as evidence in favour of his two-event model the difficulties in determining P-centres with inexperienced subjects (Marcus 1981, Cooper et al 1988); he suggests this could mean that no one perceptual event underlies subjects settings. In addition he cites the subject differences found by Cooper et al (1988) when syllable rhyme duration was varied. This he suggests indicates a psychologically weaker syllabic centre of gravity and a stronger perceptual onset. As will be described in Chapter 4, the measurement of P-centres is predicated upon subjects abilities to perform the dynamic rhythm setting task. Experiment 1 in this thesis addresses the question of whether a group of inexperienced undergraduates could perform the rhythm setting task. While some factors such as tempo affect performance, only one subject was unable to perform the task. Overall the rest of the subjects were accurate in their settings. Earlier in this chapter there was an examination of the Cooper

et al (1986, 1988) results which showed large subject differences in alignment tasks. In contrast to Pompino-Marschall, the conclusion was drawn that these differences reflected experimental problems (such as counterintuitive alignment methods) rather than robust individual differences.

P-centres, as defined and modelled by Pompino-Marschall, are thus different from the traditional approach; that of Morton et al (1976). No empirical evidence collected in this thesis was designed to explore further this central assertion that P-centres are caused by two psychoacoustic events. However the debate can become circular; P-centres are measured in rhythm setting tasks, models are developed to account for the P-centres as measured in these tasks, a failure of a model in adequately accounting for observed variance indicates a problem with the model. This is not to suggest that Pompino-Marschall is incorrect in his experimentation or his conclusions. In this thesis however the standard definition and measurement of P-centres will be used.

1.13 Summary

Acoustic signals have a perceptual moment of occurrence; this perceptual centre or P-centre is the aspect of a signal that is aligned in a perceptual rhythm setting task. In a perceptually isochronous sequence, the P-centres of the stimuli are physically isochronous. No single acoustic event of a signal (eg. vowel onset) has been found to relate to P-centre location. Various models have been proposed to account for P-centre location. Their strengths and weaknesses are described below:

Marcus's model is global and phonetically motivated, relating P-centre location to the durations of the prevocalic and syllable rhyme sections.

The positive aspects of Marcus's model are that:

1) it fits his data well

2) it is simple and testable; his division of the syllable into two portions maps onto the psychological divisions of onset and rhyme

3) it is applicable to both speech and non speech signals, as Morton et al felt such a model should be. In the case of non speech signals the 'vowel onset' is replaced by the peak increment in mid band spectral energy. This is also a benefit from his definition of vowel onset in acoustic terms.

4) it is based upon the P-centres of signals as calculated using a least sum of squares fit of all the intervals set in the experiment; the values are therefore well established, and the effect of odd or irrelevant outlying points reduced.

The arguments against his model are that:

1) His choice of the particular definition of vowel onset is quite arbitrary; there are other possible definitions, and his choice does not seem to be driven by the data

2) It is based on speech from only one speaker, and thus his model may not account for differences between speakers

3) His model locates P-centres firmly as being influenced by the durations of different sections of a syllable. It thus needs to define where the sound *begins* to determine the duration of the first section. He has for this a threshold function of the sound onset being when the signal passes -30dB SPL. This is again quite arbitrary, and means that if in continuous speech the signal never fell below -30dB SPL a new syllable with a new P-centre would not be heard. This threshold problem will affect any model of P-centre location that relies on duration of segments alone.

4) His rejection of intensity attributes as affecting P-centre location was based on one experimental finding, that doubling the amplitude of a syllable final "t" burst did not affect the P-centre of the signal. He did not investigate the effect of altering the intensity attributes at the onset of a signal.

5) Some of his experiments used stimuli which contain phonotactically illegal sounds. This may have affected his results.

6) Although he used a full P-centre setting paradigm, his lack of a common reference sound in all his experiments reduces their comparability.

Howell's model is an acoustic global model, relating P-centres to the centre of gravity of the amplitude envelope; the P-centre will thus vary with the alteration of intensity/time parameters of the syllable.

Positive aspects of Howell's model:

1) It is simple and testable

2) Aims to account for production and perception data

3) Psychologically plausible (work reviewed in the next chapter indicates that the parameters he identified affect the perception of acoustic stimuli at several levels)

4) Uses a parameter which is the principal one connect with syllables, and also applies to the whole syllable.

4) Aims to be applicable to both speech and non speech

Arguments against Howell's position:

1) Although Howell's model predicts most observed P-centre variations, it is specifically based on two sets of perceptual experimental data in which only one variable - rise time - was manipulated, and actual P-centres or analogues not calculated.

2) does not account for all the perception data - eg. lack of effect of amplified "t" burst at syllable offset (Marcus, 1981). Howell's model predicts this manipulation will have an effect due to the increased energy at the offset. Marcus found it does not shift P-centre.

3) the important parameter of the model is the intensity/time distribution of a signal, and thus it does not address any possible spectral qualities associated with P-centre location

The articulatory model of P-centres is based on patterns of articulation, perhaps related to the articulatory vowel onset, but affected by the whole syllable articulation. Adopts Marcus's onset/rhyme distinction as accounting for the observed P-centre variation and is thus overall a global model.

Positive aspects of the articulatory approach are:

1) Fits in with her general model of speech event perception and production.

2) Feels intuitively reasonable when producing rhythmic speech sequences.

Negative attributes:

1) Not initially based on any actual P-centre measures, but instead on some unsensitive measurements of produced speech and forced choice perception tests and non significant reaction time experiments. What in later experiments are called P-centre setting alignment tasks are in fact not; they simply involve setting a syllable to a rhythm with a common reference sound. The nXn matrix of trials is not performed, and often they did not balance the ordering of the stimuli. The initial A - B interval was always the same, meaning that response bias is a real problem. The use

of key presses to make adjustments of different sizes and directions is counter-intuitive compared to Marcus's potentiometer knob.

2) Specific to speech. Cannot therefore involve any of the findings from auditory perception, for example the perception of musical notes (Gordon, 1987).

3) Involves direct perception. Open therefore to refutation, if, as is the case, experiments can be carried out which manipulates the acoustic signal and thus shift the P-centres.

Other work was outlined, which suggests that P-centres are common to all languages, being a function of syllable structure, and syllables being universal.

Finally the consideration that not one P-centre, but two perceptual attributes of a signal are responsible for the physical anisochrony/perceptual isochrony observations in production and perception, was noted.

Chapter Two
Perceptual consequences of the intensity and time variation in signals

Abstract
The aim of this thesis is to model P-centres in a manner applicable to speech and nonspeech signals. This chapter addresses how the acoustic attributes of non speech sounds affect their perceptual qualities. This chapter considers evidence from physiological and psychological experiments that illuminate how amplitude characteristics such as rise time contribute perceptual experience. The implications of this research for models of P-center location will be considered.

2.1 Introduction

Howell's model of P-centre location relates the P-centre to attributes of the intensity time profile of a signal, while Marcus's model is based upon the duration of different sections of a syllable. Howell's model is conceptually related to Mermelstein's (1975) identification of the amplitude envelope as the primary acoustic cue associated with syllable structure.

The aim of this chapter is to consider evidence of the importance of intensity profile attributes in the perception of acoustic signals. Specifically the envelope characteristics that Howell's model integrates, such as rise and decay time of signals, will be addressed. Perceptual processes from the physiological level upwards will be covered. Evidence from both speech and non speech fields will be considered, since one of Howell's main points is that his model is applicable to both. The aim of this entire thesis is to model P-centres in a local, non speech specific manner. A model of P-centre location which defines the phenomena in terms of acoustic parameters can strengthen its account by establishing that such parameters are important perceptually.

2.2 Physiological Evidence

This section will briefly consider some evidence for the processing of rise time parameters by the auditory system. The relationship between physiological and psychological evidence is not always clear; this section is included to establish that such features are preserved in the encoding of information in the auditory system.

2.3 Response to rise time in the Inferior Colliculus of Bats

Suga (1971) made a study of the responses of Inferior Collicular neurones of bats to tone bursts with different rise times. He found neurones which were specialized for the analysis of amplitude modulated sound, especially for the rising phase of the amplitude change. The response pattern of these phasic on-response neurones did not change with rise time, although the response patterns of some neurones changed to inhibitory responses.

The thresholds of responses to the tones increased when the rise time was lengthened, and the degree of increase varied across the neurones. Neurones which showed a large increase in threshold were excited by a signal with a short rise time - that is, a fast increase in amplitude. Some neurones exhibited an upper threshold, above which they were not excited. This upper threshold disappeared if the presented tone had a longer rise time.

Increasing the rise time of the presented tone caused changes in the excitatory areas of the neurones. In some cases it was increased and in others decreased. This suggests that the neurones concerned are specialized for responding to tones of different rise times. Rise time changes also led to changes in the inhibitory area: the change in the inhibitory area of any neurone was not necessarily the same as that found in the excitatory area.

2.4 Response to rise time in Auditory Nerve of Cats

Delgutte (1980) investigated the responses obtained in the auditory nerve of the cat to various speech like sounds. He examined the response to fricatives, "sh" and "ch", with different rise times, and identical peak amplitudes. The "sh" had a 40ms rise time, and the "ch" a 1ms rise time. The stimuli were presented at 41dB SPL and 56dB SPL. The "ch" signal resulted in a large onset transient at both levels, the "sh" only at the 56dB SPL presentation. The "sh" response peaked before the signal rise time ended.

Delgutte and Kiang (1984 a,b,c), Delgutte (1984) found that a fricative stimulus was less well represented in the auditory nerve than a vowel. An isolated "cha" stimulus led to peak in response at the onset of the signal: this peak was attenuated by a preceding vowel or consonant. There were large adaptation effects relating to the onset characteristics of the signals. Signals with a short rise time (10ms) resulted in a peak response that was almost instantaneous with the signal onset. The response was sharply peaked. Signals with a longer rise time (75ms) led to a peak response between 15-30ms after the signal onset. The response was less sharply peaked; rather more flattened.

In conclusion it can be seen that signals with varying onset amplitude characteristics are represented differently in inferior colliculus and auditory nerve discharge patterns, in bats and cats respectively. The finding that features are preserved during certain levels of encoding does not mean that such features are perceptually salient, and the usual caveats about human/animal comparisons apply.

2.5 Perceptual evidence - discrimination of rise time

Intensity profile attributes of stimuli, specifically the stimulus rise time, are represented in the auditory system. How accurately are discriminations of these parameters perceived, and does discrimination vary with spectral content?

Van Heuven and van de Broeke (1979) attempted to establish JND's for the rise and decay times of 10Hz sinewaves and white noise bursts. The Weber fraction was a minimum at rise/decays of 80ms and increases significantly for rise/decay times below 20ms. The discrimination of noise decays was the most accurate; for the sinewaves, the discrimination was best for the onsets. They investigated this further (van den Broeke and van Heuven 1983), this time without confounding decay time and overall duration as variables. If rise and decay times are varied independently of duration, then discrimination was similar for both, and not very good in either. Their range of rise times corresponded to those found in speech sounds, ie. between 10 and 80ms. Accuracy was best for decay times in noise stimuli.

Conversely, studies have shown that subjects abilities to identify musical instruments decreased significantly if the rise time of the tones is removed - suggesting that the rise time is important and detectable, even if it can not be accurately discriminated.

This frequency/periodicity dependent attribute of rise times was found by Gjaevenes and Rimstad (1972) investigating the influence of rise time on loudness of sound pulses. The signals with the fastest onsets were perceived as the loudest. There was significant effect of the signal spectrum; the influence of rise time on loudness was stronger for signals with a central frequency of 250Hz than those of 1100Hz.

2.6 Perceptual evidence - categorical perception of rise times

Categorical perception is an aspect of perception characterized by very accurate discrimination between events in different perceptual categories and poor discrimination of events within categories. It has been of interest to speech researchers as a model for the identification of speech sounds which vary along a continuum. Given that signals, both speech and non speech, can vary along the continuum of stimulus rise time, is this reflected in a percept of a continuum, or are stimulus rise times perceived categorically?

Cutting and Rosner (1974) provided evidence that nonspeech sounds were perceived categorically. This was potentially a very important finding in terms of how unique categorical perception is to speech perception. Rosen and Howell (1981) however, identified errors in the construction of Cutting and Rosner's stimuli. Their own subjects responses were following Weber's law, in accordance with the findings of van Heuven and van de Broeke (1979). Difference limens for rise time were increasing proportionally with rise time duration. Kewly-Port and Pisoni (1984) confirmed this finding. This indicates that the rise times of stimuli form a perceptual continuum rather than two categories (eg.'plucks' and 'bows'). (Cutting (1982) replicated this, but also found that *logarithmic* increments in rise time can lead to categorical perception).

Since identification of signals constrains categorical perception, different results might be expected with musically trained subjects. Smurzynski and Houtsma (1989) replicated the categorical perception experiment with trained listeners, and found none of the discontinuities in the difference limens with rise time that would indicate categorical perception. In common with Rosen and Howell's subjects, their subjects reported that the short rise times could be discriminated by the strength of the 'thump' at the onset. This is due to spectral splatter, and

they hypothesize that a perceptual correlate of this serves as a dominant cue for discrimination.

Thus the rise times of non speech signals are not perceived categorically, although energy splatter may provide extra cues to very short rise times (that is, those shorter than 20ms).

2.7 Perceptual evidence - effect of rise time on perception of temporal order

A problem encountered by researchers who use synchronicity judgements as an experimental paradigm (Gordon 1987) is temporal order discriminations. It is possible for subjects to hear that stimuli are not synchronous, without being able to make a decision as to how they should be altered in terms of which is to early, which too late. Do attributes of the intensity profile of signals affect the ease of this task?

The first measurement of temporal order judgments was made by Hirsh (1959). The task was for subjects to decide which of two sounds was **first** in a sequence. The threshold of temporal order is determined by the difference between two onsets needed for subjects to state which sound occurs first. Hirsh found that for a click and a noise (with rise time of 15ms) sequence, where the click preceded the noise by 10ms, the subjects perceived stimuli as equally to be first. That is *there is a difference between the point of subjective simultaneity and physical simultaneity*. The subjective percept of events in time is not the same as their physical location. Thus it could be hypothesized that stimuli attributes which affect temporal order judgements, might also affect the P-centres of the stimuli. This next section will briefly consider the existing experimental evidence for this.

Hirsh considered that the temporal order judgements (TOJ) were independent of the sounds used, and thus altering frequency, intensity or bandwidth would have no effect. Conversely, he felt the duration and rise time of a stimulus might affect the TOJs.

His evidence supported the latter claim. If the stimuli had very rapid rise times (2ms) the subjects were more sensitive to differences in onset times, and this was reflected in their TOJs. Subjects made more accurate judgements of temporal order if the stimuli had rapid rise times, suggesting that this made the stimuli easier to discriminate. In click/tone sequences, the click had to precede the tone by increasing amounts, as the tone duration was increased, for the stimuli to be perceived as simultaneous. That is, the threshold is increased. Thus Hirsh's experiments show that duration and rise time affect the discrimination and perceptual synchronicity of auditory stimuli.

Pastore, Harris and Kaplan (1982) replicated these experiments and found that temporal order thresholds (for TOJs) were direct functions of the stimulus duration and rise time (for longer rise times, ie. bigger than 10ms). Pastore (1983) explicitly compared the effect of rise time to Vos and Rasch's (1981) work on perceptual onset. He extended the 1982 paper to offset asynchronies - the TOJ for offsets. He found that the duration stimuli affects the TOJs in the same way as for onset asynchronies. The superiority of offset order thresholds was accounted for by the possible use of echoic memory.

2.8 Stimulus rise time and musical tones

The rise time of signals therefore affects various aspects of how the signals are perceived at several levels. How does this dependency relate to the timing of signals in a rhythmic context? That is, how do the rise time of stimuli affect their perception in a musical context?

In a paper reviewing the psychophysical attributes of musical notes Terhardt (1978) stated that "..perceived equability is not identical with equal time intervals between physical onsets" (1978, p484). The observed deviation from isochrony, he found, was due to the duration and amplitude envelope shape of the signals concerned, but independent to tempo, sensation level and spectrum (Terhardt and Schütte 1976). Obvious parallels between the observations that led to the development of concept of P-centres can be seen here. Physical isochrony does not necessarily result in perceptual isochrony. Attributes of the signal intensity/time distribution affect the systematic deviations which lead to perceptual regularity. This parallel phenomena will be considered in more detail in this section.

Research by Schütte (1977, 1978) on this topic related the perceptual onset (P.O.) of tones to a first-order leaky integrator circuit, characterized by a time constant τ. The inputs were tones and the outputs were subjective envelopes (Schütte 1977, 1978). In this model, the perceptual onset was the moment when the subjective envelope passed a threshold, relative to the maximum value of the envelope. The threshold was variable relative to the physical intensity profile and duration. Efron (1970a, 1970b, 1970c) however found that perceptual onset was independent of physical duration.

2.9 Work of Vos and Rasch

Rasch (1979) investigated synchronization in performed ensemble music. He observed that in polyphonic music, the synchrony dictated by the score is never realised in the actual performance. In fact, the standard deviation of the differences of onset time in trios (three instruments) was typically 30-50ms. Nevertheless, this does not generally lead to the percept of uneven playing. There are similarities in the conceptual motivation for his study with that of Morton, Marcus and Frankish (1976); Morton et al wished to define what is

evenly timed in a perceptually isochronous sequence, if it is not the physical onset; Rasch wished to define what is synchronous in a polyphonic piece, if it is not the physical onsets of the instruments?

He studied the onset difference time of trios; that is the measured difference between physical onsets of instruments. The onset time was defined as when the intensity passed a threshold of 15-20 dB below maximum.

He compared the onset difference times of six pieces, each involving different instrument combinations.

Piece 1:	three recorders (treble and tenor)
Piece 2:	three recorders (treble, tenor and descant)
Piece 3&4:	oboe, clarinet and bassoon
Piece 5&6:	violin, viola and cello

The mean asynchronies (ie. standard deviations of onset different times) were 30ms and 31ms for the two recorder pieces, 37ms and 27ms for the wind instrument pieces, and 49ms and 37ms for the string pieces. Thus the range of asynchronies varied with the different trios.

When the mean physical onset differences were considered, it appeared that the tenor (low frequency range) was most different from the treble in the first piece, and from the descant in the second piece; they seem to vary thus with the frequency range of the instrument. In the wind pieces, the clarinet was most different from the oboe. In both the string pieces the viola was most different to the violin.

Rasch felt that possibly the observed differences were related to the rise times of the notes of the instruments. He stated that the recorders all have short rise

times ".....so the *beginnings are clearly marked*" (Rasch 1979, my italics). He stated that the wind instruments also have relatively short rise times. The string instruments have longer rise times - between 50 and 200ms. Rasch hypothesized that shorter, sharper rise times make better synchronization both necessary and possible.

It must be stressed that the emphasis of this paper is less about the perceptual beats of the musical notes, and the implication for performed and perceived synchronization, than about how longer rise times lead to a failure of performers and listeners to notice asynchronies.

This concept was refined and extended in a later paper. Vos and Rasch (1981) studied the Perceptual Onset (P.O.) of musical tones. This they defined within all acoustic stimuli, as "the moment in time at which the stimulus is first perceived" (1981, p323). This opposed to the physical onset, which is the time when the signal is first generated. The perceptual onset is delayed relative to the physical onset. Note again that this concept of perceptual onset is again similar in motivation to the P-centre hypothesis, which they explicitly mention.

Vos and Rasch (1981) attempted to apply a simple threshold model to the perceptual onset of musical tones. To test this they used a dynamic rhythm setting task, and their experimental method made similar assumptions to those of Morton et al; tone sequences were defined as perceptually isochronous if the time intervals between successive perceptual onsets are equal. In a method similar to Marcus (see Chapter 4) this enabled them to determine the threshold amplitude for perceptual onset.

In this way they measured the effect of rise time and intensity of the stimuli, and the effect on perceptual onset. An important difference between this method and that of Marcus, etc, is that to alter the A - B interval, the subjects

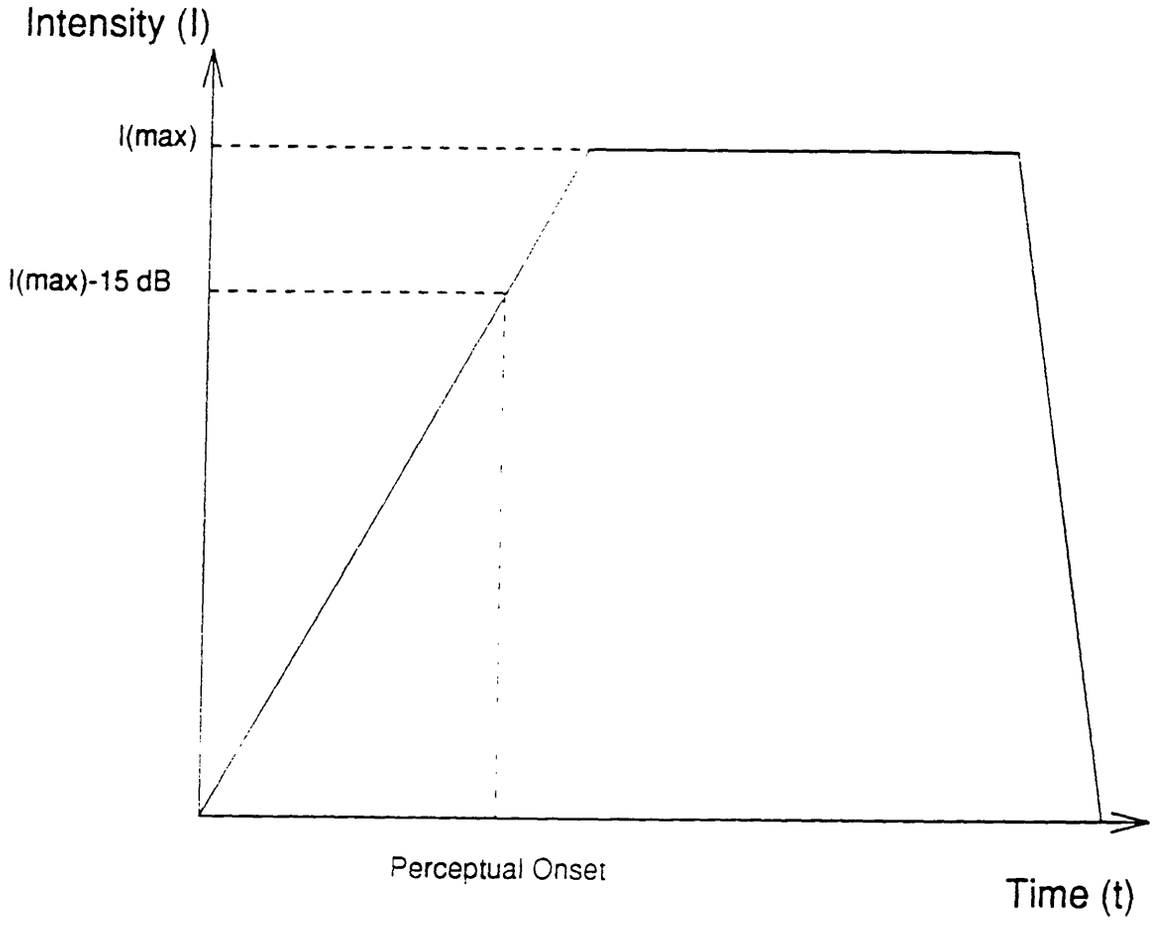
in Vos and Rasch's experiments had to vary the duration of the B stimulus. That is, to shorten the A - B interval, they increased the steady state portion of B.

Their results indicated that the P.O. of a signal was located at the point that the physical onset passed 17% of the maximum - that is, 15dB SPL below maximum intensity.

They next varied the physical intensity of the signals. They found that the time difference between physical and perceptual onsets *increases* with *decreasing* tone intensity. This increase was described as a shift up of the relative threshold which leads to the perceptual onset; they stressed that this shift in threshold was small compared to the variation in intensity. This led them to conclude that the threshold could simply be define relative to the maximum intensity of the signal, regardless of the physical intensity.

Their final experiment investigated the signal to noise ratio of the experimental presentation and its effect on the threshold. They found that the relative threshold for the perceptual onset of the tones decreased with increased level above the masked threshold. These results were congruent with the findings of the previous experiment - the louder the tones, the smaller the threshold.

They thus concluded that the P.O. of a tone was due to the intensity of a signal passing a simple threshold. The threshold was relative to the maximum intensity of the signal, 6-15dB SPL below the maximum intensity of the tone, depending on the sensation level of the tone. Figure 2.1 shows a diagram of this model of perceptual onset. Their results seemed to suggest adaptation to the constant stimulus level, and indeed the repetitious nature of all dynamic rhythm setting tasks are "optimal conditions for adaptation" (1981, p331). They found that the temporal integration model (Schütte 1977, 1978) predicted the



Vos and Rasch's (1981) model of perceptual onset (defined as P-centres) defines the defines perceptual onsets as the time that the envelope passes a relative threshold of 15dB SPL below the maximum intensity.

Figure 2.1 Vos and Rasch model of Perceptual Onset (analogous to Perceptual Centres) 1981

general trend of the results, their threshold model was more powerful. Marcus (1976) had discounted a simple temporal integration model as a determinant of P-centre location. A psychological explanation of their model was that "the adaptation of the hearing mechanism to a certain stimulus level is responsible for perceptual onset" (p334).

Since this model is dependent upon the maximum intensity of a signal, it is not congruent with some of the speech experiments mentioned earlier; specifically Marcus's finding that increasing the intensity of a syllable final "t" burst has no effect on P-centre location. If the "t" burst amplification altered the maximum amplitude of the entire signal, Vos and Rasch's model would predict an alteration in the P.O. since it is reliant on the maximum intensity to determine the threshold.

2.10 Seton's work - an extension of Vos and Rasch

Seton (1989) replicated Vos and Rasch's experiments, with both speech and non-speech sounds, and using subjects who were not accomplished musicians. He found that stimulus pairings of two stimuli with short rise times gave lower onset thresholds, than pairings with longer rise times. Vos and Rasch did not find this, which was part of their claim that for a fixed peak level and background noise level, the onset threshold was a constant.

Seton also found that his vowel stimuli did not show a rise time dependent threshold shift. This, he concluded, was due to the duration and frequency of the glottal pulse. The duration of the pulse (8ms) was long compared to the rise times used, and the sharp rise/decay in amplitude over the pulse cycle may have led to the second glottal pulse becoming the effective onset.

Next, Seton repeated this experiment with tone stimuli only, and only presented stimuli whose rise times were different by 20ms. Thus pairings such as 20+40ms, 30+50ms were made. He found, as in the previous experiment, that for relatively short rise time pairings, the onset threshold is at a lower level than for the longer rise time pairings. He examined his short rise time stimuli for transient clicks and 'glitches', and found none. He concluded that rapid onsets might be processed differently to longer onsets. He also found a similar pattern of offset thresholds for stimuli in which the offsets of were altered in decay time, although the effect looks smaller than the onset manipulation results.

Seton concluded that his experimental findings could be used to modify Vos and Rasch's model, to account for the short/long rise time effect on the thresholds. He put the boundary between "fast" and "slow" onset events between 30 and 50ms.

2.11 Further work on musical notes - Gordon's model

Gordon (1987) conducted a series of experiments to measure and model the perceptual moment of musical attack - perceptual attack time (PAT). He defined this as the moment "a tone's moment of attack or *rhythmic emphasis* is perceived"(1987 p88, my italics). Gordon explicitly stated that this was not the same as the perceptual onset time. Although the two could coincide, there are instances where they do not (eg. a bowed instrument has a clear onset, followed by a perceptual beat). He adds that this a distinction not made by Vos and Rasch, who used perceptual onset time to mean PAT as he defines it. The PAT of a musical note is also related to the P-centres of speech sounds. Gordon concludes that since there seems to be an 'inverse' relationship between rise time and PAT (shorter rise time, clearer PAT), the rise time of a tone is a reasonable place to start.

He tested this hypothesis using two experimental methods. In the first, judgements were made about stimuli isochrony (cf. dynamic rhythm setting task). The other two experiments used judgements of synchrony.

He used 16 instrumental tones, all at the pitch of $E^{\flat}$ above middle C (about 311Hz). He used a reference sound in the isochrony task - all judgements were made against the $E^{\flat}$ clarinet. In the first synchrony task, three standards were used, the clarinet, the cello and the bassoon. The clarinet has 'average' attack qualities, the bassoon a very quick attack, and the cello a long, drawn out attack. In the second synchrony task a drum sound was used as a reference.

The isochrony task was a standard rhythm setting task, as used by Marcus (1981) Cooper Whalen and Fowler (1986, 1988). A sequence of alternating tones is presented to a subject, who can alter the intervals until they are satisfied that the sequence sounds even. The subjects in this experiment controlled the intervals with a slider (as opposed to a potentiometer knob (Marcus 1981) or a series of keys (Cooper et al 1986, 1988). In each sequence a stimulus was set against a common reference sound. All stimuli were set to an even rhythm in this way. The A-A interval was 1200ms, the initial A-B interval was quasi-random (slider randomly put at either extreme).

In the synchrony tasks, the subjects were required to synchronize the output of two stimuli, as if they were trying to get two players to perform exactly together on the beat. The difference in PAT across instruments (ΔPAT) was measured by taking the offset from isochrony in the first experiment, and the measured delay (in ms) between the physical onsets in the second experiment.

Analysis of the results showed that the relative PATs for the sixteen instrument stimuli were different. In the experiment with three reference sounds, the effect

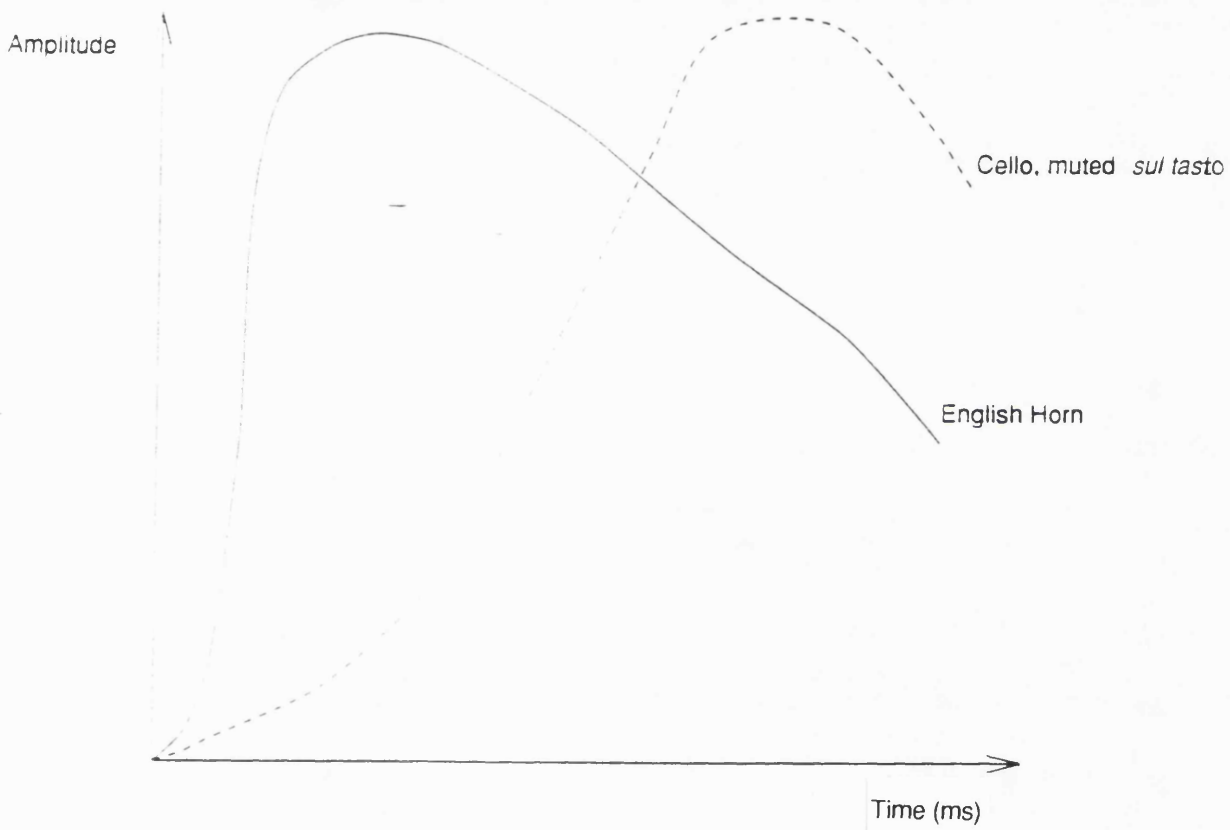
of the reference sound was significant. This was expected, since the three reference sounds were so different in attack attributes.

There was a significant effect of subject on all the results; this was only regarded as a cause for concern in the last experiment(synchronization with a drum beat). In the other experiments the subjects were following similar patterns but covering slightly different ranges (smaller than those noted by Cooper Whalen and Fowler 1986 for speech). Against the drum, the synchronization settings are very varied. Against the bassoon, they were much less so.

Problems with the synchronization task (which was expected to produce more accurate results than the rhythm setting task) were deemed as result of difficulties with temporal order and fusion. It was easier for subjects to detect that tones were not synchronous than decide how to adjust them to reduce this. Thus there was a problem with detection compared to order judgements. Subjects also experienced fusion phenomena - vertical fusion (individual timbres not being resolved from the mixture) and horizontal (two events being perceived as one). This was worst for the string tones.

Overall the tasks the subjects found hardest were those against the drum sound. This was mainly due to the temporal acuity mentioned above, and seemed to be precipitated by the very sharp attack time of the drum beat (10ms). Alternatively masking may have contributed to this, the drum attack masking the tone onset attributes.

Generally the results of the rhythm setting task corresponded with the first synchronization task. Where there were discrepancies (some of the string instruments and the flute), all have gradual, steady increases in amplitude at the onset, and sometimes spectral changes occurring after the onset (flute,



These slopes represent the initial portion of the amplitude envelopes of two of Gordon's 1987 stimuli. The Perceptual Attack Times (PAT, defined as P-centres) of these two instruments vary; The horn has an immediate PAT due to its rapid initial increase in amplitude; the cello has a later PAT due to its more gradual initial amplitude increase.

Figure 2.2 Gordon's model of Perceptual Attack Time (defined as Perceptual Centres), demonstrating the difference between the onset amplitude characteristics of the cello and the english horn.

trombone). These may be cues to the PAT. He concludes that "PATs for tones with impulsive attacks seem to coincide with he physical time of attack regardless of whether presented isochronously or synchronously with other tones; however PATs for tones with nonimpulsive attacks depend on the mode of presentation" (1987, p100). That is, the PAT of signals is determined by rise time characteristics and the experimental task.

When modelling these results, Gordon found that the Vos and Rasch threshold needed to be reduced from 17% to 5.8% of maximum to fit the data well. Instead, a model based on a slope threshold normalized with the rise time, fitted the data well - if power slope envelopes were used (as opposed to amplitude envelopes).

Gordon concluded that when a tone's rise time is rapid, its PAT is determined by amplitude cues; when the rise time is long, the PAT is affected by spectral attributes. When the rise time is longer, the PAT is less distinct. Gordon relates the PAT to short term adaptation in the firing of auditory fibres; both phenomena are related to the sensation level and rise time of the stimuli. This suggests the slope of the signal rise function as a prime determinant - Figure 2.2 shows a diagram of this.

2.12 Summary

Intensity time attributes of non speech and speech signals affect perception at many levels of analysis, from encoding in the auditory system, through identification and discrimination and judgements of temporal order. They were finally shown to be important parameters in the perception of the isochrony of sequences of musical notes. Such parameters have been identified as important in determining P-centre location (Howell 1984, 1988a&b) and discounted as determinants by several researchers (Marcus 1981, Tuller and

Fowler 1980). There is thus an incongruity between the importance of intensity attributes such as rise time and duration on perceptual regularity in musical sequences, and some of the reported effects in speech sequences. These incongruities will be addressed in more detail in the next chapters of this thesis.

Chapter Three
Outline of Thesis

Abstract
The aims of this thesis are: to examine three current approaches to P-center modelling, and describe how well each accounts for P-centre locations in different experiments: to attempt to model P-centres from a local, non speech specific perspective. This chapter will outline the aims of the different experiments in the thesis, and the implications of the experimental results.

3.1 Introduction

As has been described in the two previous chapters, the existing research on P-centres in speech and the PAT of musical tones suggests a role for amplitude varaition as a determinant for both phenomena (Howell 1984, 1988a&b, Gordon 1987, Terhardt 1978, Vos and Rasch 1981). In addition, time/amplitude parameters such as rise time and duration have been shown to affect perceptual judgements about stimuli, for instance in temporal order judgment thresholds (eg. Pastore 1982). Such studies further stress the role these parameters play in the perception of acoustic events. The converse view to this is that P-centre location relates to articulatory gestures in the production of speech; that these gestures are directly perceived, and acoustic attributes play no role as determinants (Fowler 1979, 1983).

In this thesis, P-centres will be studied within the same conceptual framework as musical tones, and other non-speech events. A model will be developed which aims to account for experimental findings in both speech and non speech research. The model will thus be applicable to any discrete acoustic event, and not unique to speech. Arguments against this approach would state that speech is perceived specially, and cannot therefore be compared to non-speech stimuli

in terms of how the signal is processed by the ear. In addition, many researchers who take this view would consider that speech is not only perceived specially but directly, without any processing of the information. Thus any approach that considers the acoustic attributes and their effect on perception is thus considered misguided.

Arguments in favour of the proposed approach are that it will force an explanation that is in accordance with models of general auditory perception, and avoid the short-cuts inherent in considering them as specific to speech. Further arguments in favour are the many experimental results that indicate that parallels between speech and non-speech stimuli do exist. The hypothesis that P-centres are not speech specific should be tested experimentally, before any such assumption could be drawn.

A second aim of the thesis is to resolve the disparity between the speech based accounts of P-centre location, and the non-speech models of perceptual onset/attack. As outlined in Chapters One and Two, models of P-centre location in speech signals are **global**; they relate P-centre to the entire syllable structure, rather than to a single acoustic event. The models of musical 'perceptual attack' or 'onset' are **local** models; they relate the perceptual moments of occurrence to discrete, punctate acoustic events. Both the model of Vos and Rasch (1981) and Gordon (1987) relate the perceptual moment of occurrence to onset amplitude events (characterized differently in each model); their models do not incorporate duration information or offset amplitude events.

The principle problem with any global model of P-centre location is that they require the entire signal to be encoded before the sound is heard to happen. Whilst this might be possible within the repetitious dynamic rhythm setting task format, it is difficult to use such models to account for the perception of rhythmic patterns in unknown speech sequences, such as poetry. Global

models of P-centre location also require the speaker of a rhythmic sequences to use information about aspects of the entire signal they produce to time their utterances. This is a heavy computational task, and does not accord with subjective accounts of rhythmic speaking.

The aims of this thesis are therefore to test current models of P-centre location, to see which best account for the experimental results; to unpack the hypothesis that amplitude/time parameters are important determinants. Specific parameters such as the duration, rise time, quality of amplitude/time variation, and spectral content of a signal will be considered. These parameters have been shown to affect rhythmic centres and segmentation in acoustic signals (Howell 1984, Marcus 1981, Vos and Rasch 1981); to use the results of these experiments to direct further specific experiments: the data collected will used to develop a model of P-centre location. This will aim to model acoustically the patterns of rhythmic centres in speech and non speech sequences; the model will aim to be local, rather than global, and be congruent with work from the non speech domain.

3.2 Experiments

3.2.1 Production/perception of rhythmic sequences

Research on speaker production of rhythmic sequences of syllables in time to a metronome have indicated consistently that speakers will physically mistime their productions to synchronize their utterances with the metronome beats. Syllables with long consonant clusters at the onset are reliably produced earlier with respect to the metronome beat, than syllables with shorter clusters, or no consonant at all. Such findings have typically been related to the P-centres of syllables, where a syllable with a longer consonant cluster would be found to have a later P-centre. It has thus been accepted by theorists that P-centres govern both the production and perception of rhythmic speech sequences. In

this experiment, rhythmic sequences are produced by seven speakers without a metronome beat present. This is to encourage participants to produce perceptually isochronous sequences without the external rhythmic context of a metronome. Sequences judged to be rhythmic by both the speaker and two judges were analyzed. The pattern of onset to onset intervals found in production are compared to those found when samples of each participants' utterances are set to a rhythm together in a perceptual rhythm setting experiment. No current models explicitly predict that there will be any difference between perception and production. If P-centres govern the production of internally generated rhythmic sequences and perception in a rhythm setting task, then the pattern of intervals set will be similar to those produced. If they do not, then a breakdown between the perception and production of rhythmic sequences is implied. The ratio of A-B:B-A intervals for perception and production sequences for all speakers will be plotted to investigate the slope of the relationship.

3.2.2 Speaker Differences

Most experimental perceptual P-centre settings involve speech from only one speaker. Do different speakers produce speech with different acoustic and articulatory characteristics, leading to different rhythmic centres in the speech signal? Experiment 2 addresses this; speech from eight different speakers is used, the P-centres are determined in a rhythm setting task, and related to the acoustic profiles of the signals.

An **amplitude/time model** (eg syllabic centre of gravity, Howell 1988 a&b) would predict that differences might well arise between speakers; if the speakers varied in their speech in a way that affected their amplitude variation, then their P-centres would vary.

The **phonetic model** (Marcus 1981) would predict differences between speakers only if the differences led to varied durations of pre and post vocalic

segments of the speech signal. Variations due to amplitude change would not necessarily affect the P-centre.

An **articulatory model** (Fowler 1979, Cooper, Whalen and Fowler 1988) would not predict differences in the sense that the model is wholly descriptive; it will account for any experimental findings. It cannot make any experimental predictions. It is therefore not falsifiable.

If the different speakers' syllables display different P-centres, and the Marcus model can be generalized to different speakers, then the predictions his model makes should predict the P-centre locations. If not, then it suggests that his model cannot generalize.

3.2.3 Amplitude/Time distribution

Much research in the field has indicated an effect of the amplitude profile as a determinant of P-centre location. Other theories have precluded such parameters. The next two experiments try to reproduce the findings of two such researchers, since they disagree to such a marked extent with existing amplitude/time model results.

3.2.3.1 Researchers (Tuller and Fowler 1981), working from an articulatory perspective ran experiments which aimed to provide proof for their theory that the time/amplitude profile plays no role in P-centre location. They found that removing all time/amplitude variation from a speech signal did not affect its P-centre location. They concluded that underlying acoustic gestures were responsible.

An attempt was made to replicate this experiment since the results were incongruous with the time/amplitude research mentioned earlier; extra care was taken to avoid the practical problems which marred the original research (see Chapter 1 for details).

The **acoustic, syllabic centre of gravity model** (Howell 1988 a&b) would predict that removing all the amplitude/time variation from a signal would alter its P-centre, since it would dramatically alter the amplitude distribution in the signal.

The **phonetic model** (Marcus 1981) would not explicitly predict that this manipulation would cause the P-centre to change, since the duration of the signal would remain the same. However, if the manipulation altered the signal such that it affected the acoustic definition of vowel onset, it would affect the segmental durations, and thus the P-centre location.

The **articulatory model**, as defined in this case (Tuller and Fowler 1981), would predict that the P-centre would not be shifted, since the pattern of underlying articulatory gestures would be unchanged.

If the infinite peak clipping does not shift the P-centre of the signals, then this replicates the Tuller and Fowler finding, and discounts the amplitude/time profile as a determinant of P-centre location. Furthermore, it suggests that since there is a lot of evidence from non-speech material of the importance of amplitude/time distribution, that speech is indeed perceived differently from such stimuli. If the infinite peak clipping does affect the P-centre location, then this indicates that the amplitude time profile of a signal does affect the P-centre location, in accordance with research on non-speech stimuli.

3.2.3.2 Final "t" burst amplification. In his thesis, Marcus (1981) conducted several experiments in which he manipulated speech waveforms to observe the effect on P-centre location. The majority of these involved the variation of the durations of the segments; in fact he only made one explicit manipulation of the amplitude of a speech sound. This was an amplification of the syllable final "t" burst in the word "eight". He found that this did not affect P-centre location. Increasing the length of silence in the closure before the burst did change the location, however. From these results he developed his model in which the duration of segments of a speech sound contribute differently to the P-centre location, and the amplitude variation plays no role.

This experiment will replicate his "t" burst study in a perception experiment. It will extend the study by examining the effect on the timing of produced rhythmic sequences when subjects have alternately to say "eight" with an amplified "t" burst.

The **acoustic model** (Howell 1988 a&b) would predict that this manipulation does affect the P-centres in both perception and production. This is because the increase in acoustic energy caused by the "t" amplification would shift the P-centre back in time towards the offset of the syllable.

The **phonetic model** (Marcus 1981) obviously, predicts that a replication of this experiment would show no effect of "t" burst amplification on P-centre location, in perception and production.

The **articulatory model** (Fowler 1979, Cooper et al 1988), as mentioned earlier, would not make specific experimental predictions; however some researchers in the area have linked P-centre location to an articulatory vowel onset gesture. Thus a case could be made that this manipulation would not affect the P-centre in perception or production, since it occurs after the vowel onset.

If this manipulation shifts the P-centre, then this is evidence for Howell's model. If it does not, then this evidence either in favour of Marcus's model or an articulatory account. The conclusion could also be that amplitude changes at the offset do not affect the P-centre of a signal.

Model	**Speaker Differences**	**Infinite Peak Clipping**	**Final "t" Amplification**
Acoustic	Shift in PC if amplitude envelope alters	Shift in PC	Shift in PC
Phonetic	No shift	No shift	No shift
Articulatory	Shift in PC if articulation alters	No shift	No shift

Table 3.1 Summary of experimental manipulations, and prediction from models of whether this would affect the P-centre (PC)

3.2.4 Modelling Experiments

The outcome of these experiments highlighted problems with all the current models of P-centre location. A new model would have to consider the amplitude profile of a signal; there might be frequency specific aspects of the amplitude distribution; and events at the onset of the sound were more important than those at the offset. In an attempt to derive a new model therefore, a series of experiments were carried out. In these experiments, properties of the speech sounds were systematically varied, whilst others were held constant. This was to parameterize the proposed model.

When developing his model, Marcus (1981) performed no explicit manipulations of acoustic properties such as rise times or steady state vowel duration, to examine their effect on P-centre location. The majority of his experimental variables were dictated by existing properties of his stimuli. This may have caused errors in the resulting model.

In the following experiments, systematic variations to both naturally produced and synthesized speech stimuli were carried out.

3.2.4.1 The effect of **stimulus rise-time** on P-centre location was examined. Previous work (Howell 1984, Gordon 1987, Terhardt 1978, Vos and Rasch 1981) had indicated that this parameter does affect the beat location of a signal, whether speech or non speech. To extend these studies to speech signals, experiments were carried out in which the rise times of syllables were manipulated. In addition the possibility that not all speech segments affect P-centres was tested by varying the rise times of onset phonemes that have different spectral contents. That is, to test whether there are frequency specific P-centre changes. Any frequency dependency could relate to the perceptual strengths of different phonemes. Three sets of stimuli were created: Vowel onset ramped; CV clusters onset ramped, where C is a semivowel; CV clusters onset ramped, where C is an affricate. This was to investigate whether the effects of ramping are linear, and whether the frequency content of the ramped portion determines the effect of the ramping on the P-centre.

Chapter Four
Methodology - how to determine P-centers in perceptual tasks

Abstract
How can one measure experimentally precisely "when" in time a subject hears a sound "happen"? This Chapter will consider different experimental methods, and outline a paradigm for an *indirect* measure of P-centre location (the dynamic rhythm setting task). An algorithm for calculating P-centres from experimental results will be expressed. An experiment is described which tests some of the assumptions underlying this paradigm, and the General Method used throughout this thesis will be stated.

4.1 Introduction

The experimental measurement of P-centres suffers from a methodological problem; how can one measure where in time a participant hears a sound happen? The two solutions to this problem differ in that one attempts to measure directly where in time the participant hears the sound. The second uses an indirect measure, from which inferences can be drawn about where the participant experiences the sound.

4.2 Direct measures of P-centre location

These broadly correspond to motor tasks. The basic assumption is that if a participant is asked to tap along to a sequence of sounds, then the temporal positions of the taps can be compared to corresponding points in the signal. This will indicate where in the signal the participant hears the beat.

4.2.1 Tapping tasks

Initial experiments using this paradigm involved participants tapping along to whole sentences, either contemporaneously, or retrospectively (eg Allen 1972). The assumption here was that subjects tapped on the perceived beats of syllables in sentences, and thus that a comparison of where the taps are in relation to the speech signal would reveal the acoustic locations of rhythmic centres in the sentences.

The contemporaneous taps involve a large predictive element in order for the tap to co-occur with the signal; thus the assumption cannot be made that the participant is tapping when they hear the sound. Retrospective tapping is contaminated by encoding effects; participants tend to regularize the repeated taps, thus reducing any initial anisochrony. In addition, a finding which affects both types of tapping task is that when participants are asked to produce random taps, their inter-tap intervals are still simple multiples and sub divisions of each other (Martin 1972). This corresponds to work on spontaneous tapping, and is an indication of how geared we are to producing regular motor sequences.

A way of avoiding the problems of anticipation and encoding/reproduction effects that occur when participants tap along to a sentence would be to use just one word/signal at a time. This sound could be repeated isochronously, leading to a regular rhythm. Participants asked to tap along to such a sequence would still have to make anticipatory gestures to tap on the beat, but since the rhythm is regular and they know this, the anticipation would be the same for each repetition. Any possible source of error would be distributed over all the trials. The time of the taps and the corresponding point of the signal would then be compared, to determine the moment of occurrence, or beat of the signal. This method would also appear to increase the accuracy of the measure, since

the participant would have sufficient presentations of the experimental stimulus to be careful in their chosen taps.

There is a large body of research which uses just such a paradigm. These experiments are concerned with how participants synchronize their motor actions with environmental sounds, for example clapping along to a sound. The results of these experiments show a consistent deviation from synchrony when participants tap along to a regular sequence. They tap always just before the signal begins. The participants report that this is indeed the point in time which feels most synchronous with the sound. This fits into the areas of research which consider this perceptual/motor synchronization disparity (eg. Auxiette and Gerard 1992, Prinz 1992, Semjen, Schulze and Vorberg 1992), but also ensures that the tapping along task cannot be used to determine the P-centre of signals.

All the tapping studies suffer from not directly gaining access to the listeners' perception of isochrony. Such tasks do not control for the effects of the subjects' encoding and representation of rhythm (Seton 1989). They rely heavily on a subjects' memory for a rhythm, and ability to reproduce it.

4.2.2 Reaction time tasks

A different style of direct measurement of the moment of occurrence of a signal and thus the P-centre location, could be developed by using a reaction time paradigm. Parameters such as stimulus rise time have been shown to affect reaction times (Seton 1989). It might be hypothesized that this measure be employed to give an indication of P-centre location. However it could be just as well argued that this paradigm encourages participants to attend to the perceptual onset, rather than beat, of the signal. Thus any effect of rise time

would be more likely to be due to the signal passing a perceptual threshold sooner or earlier and causing a shift in reaction time.

4.2.3 Synchronization tasks

The final paradigm of P-centre location does not involve a motor task of any kind. It is instead perceptual; participants are required to synchronize the output of two signals (Gordon 1987). It is assumed that they will synchronize the P-centres rather than the onsets. Thus if one short burst of noise is used with other longer signals such as speech segments, the point of perceptual synchrony could be analyzed in terms of the physical onsets of the two signals. The bigger the disparity, the more different the two P-centres, and it could further be suggested that the absolute size of the difference, in ms, gave a measure of the P-centre location for the longer sound.

The problem with such a method is the difficulty of the task. The signals mask one another leading to inaccuracies in the synchrony settings. Temporal order discriminations become difficult in these conditions, and subjects can often hear that the two stimuli are not isochronous without being able to make the correct adjustments to lead to isochrony (Hirsh 1959).

A final problem for all direct measures of P-centre location is that the absolute P-centre is not known for any signal. Thus any direct measure of P-centre location which depends on the synchronous comparison of two signals is impossible, since no signal could be used as the baseline against which other signals would be compared.

4.3 Indirect measures of P-centre location

A direct method of measuring when in time people hear a sound happen is thus not possible. Indirect measures must therefore be used. For example, the perceptual occurrence could be determined by getting participants to set sounds to a rhythm. It would be assumed that they would align the signals to a rhythm based upon the 'beats' of the signals. Therefore the intervals that they adjusted when setting the stimuli to a rhythm could be used to calculate the beat locations of the stimuli. This calculation would not be possible if the sequence set to a rhythm consisted of only one sound. Instead two different signals are set to a rhythm. If they have different P-centres, then the pattern of physical intervals which lead to perceptual isochrony will be anisochronous. The P-centres are thus determined by calculating back from the set intervals.

Such a method is called *a dynamic rhythm setting task*. This involves a subject altering the physical intervals between two repeating sounds, until perceptual isochrony is achieved. The final intervals chosen are recorded, and are used to determine a 'P-centre fit' for each sound used in the experiment.

Variations of this paradigm have been used by several researchers. All have tried to measure the moment of occurrence of a sound by seeing how that sound is set to a rhythm. The precise details vary, however. Vos and Rasch (1981) had a paradigm where the duration of one sound was varied until its beat relative to another was perceptually even.

In the field of speech research there are several variations on the rhythm setting theme. Pompino-Marschall (1989, 1991) used a sound set to a rhythm against a brief transient click. He made the explicit assumption that the positioning in time of this click relative to the signal gave a direct measure of the P-centre of the signal. Cooper, Whalen and Fowler (1986, 1988) used a

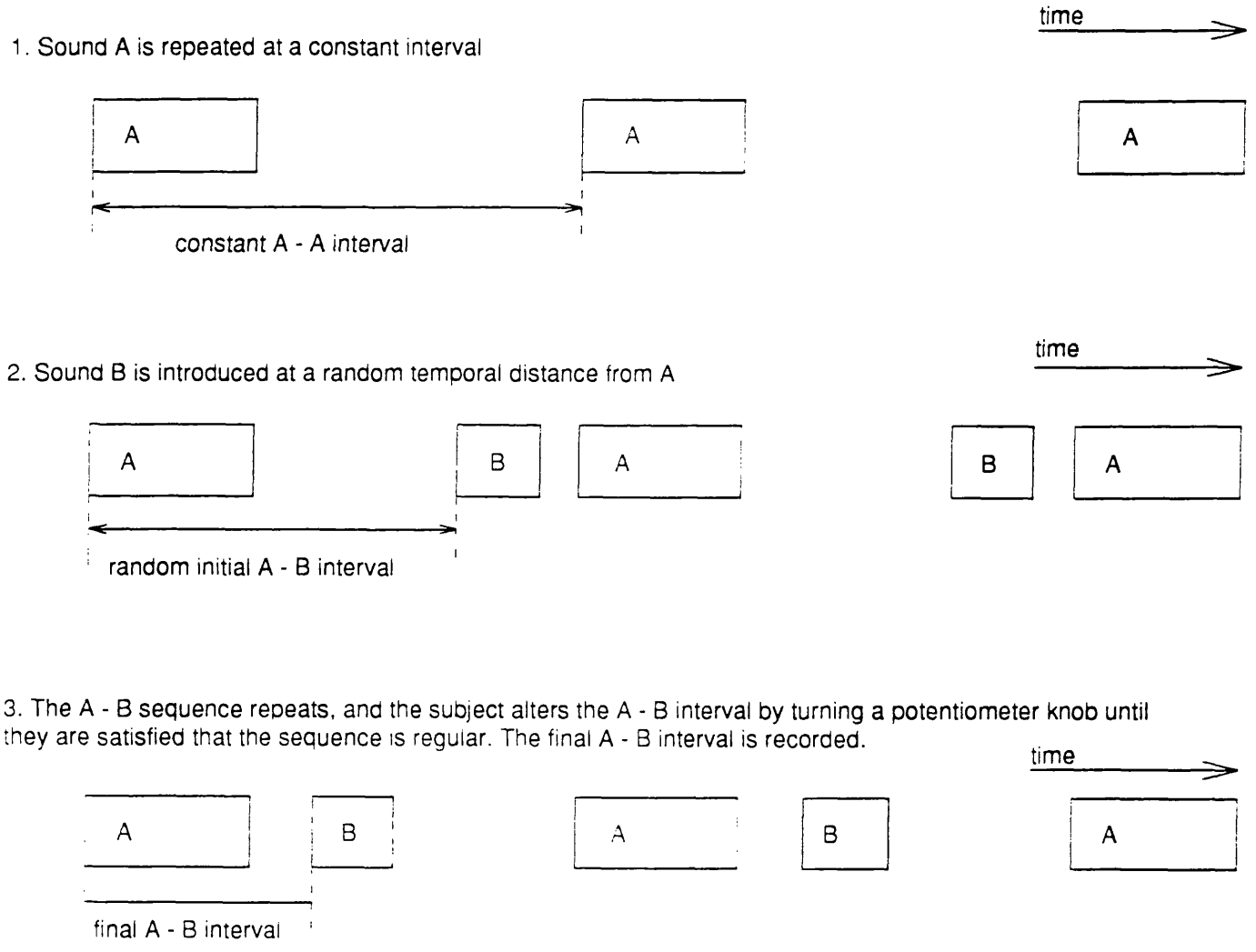


Figure 4.1 P-centre alignment task

speech sound set to a rhythm with a reference speech sound, usually "ba". They use the deviation from isochrony that is observed as an indication of the P-centre of the experimental stimulus - the bigger the deviation, the later the P-centre.

4.4 Dynamic rhythm setting task

The method originally developed by Marcus to measure P-centres by Marcus involved a large number of experimental trials. This increased the experimental degrees of freedom. All the stimuli used in a single experiment were set in pairs against each other in dynamic rhythm setting tasks. Since every stimulus is aligned to a rhythm against every other stimulus, this method provided information about how all the stimuli are set in different contexts. All the intervals were used to make a **P-centre fit** using an algorithm described below. Marcus's method is the method used in this thesis. Although it requires many more experimental trials than other methods, it produces concrete P-centre values which can then be compared to model predictions. All other methods result in adjustment values for stimuli set in pairs, which cannot always be explicitly compared to P-centres.

4.5 General design

The experimental set-up for determining P-centre location involves a judgment of 'rhythmicity' by the subject. Sound A is presented to a subject, at a constant (A - A) interval. This sound is repeated at this constant rate several times, then a second sound, B is inserted between A - A, so that the sequence is now A - B - A - B - A etc.

The A - B interval is altered by the subject (usually by turning a potentiometer knob which is finely callibrated), until the A - B - A - B sequence sounds

regularly spaced in time, that is, that the intervals between the sounds appear equal. The subject then indicates that they are happy with the evenness of the intervals, the trial is halted, and the A - A and A - B intervals are noted. Figure 4.1 shows a diagram of the alignment process.

Roughly, a deviation of the final positioning of B from **exactly halfway** between A and A indicates a difference between the P-centres of A and B; Marcus (1981) devised an algorithm for calculating the relative P centre locations for the sounds in a trial. The matrix is then folded to combine the two possible orderings for each pair of stimuli, the folded matrix is inverted, and the best P-centre estimates are fitted for each stimulus. It must be emphasized that the P-centre fits, calculated in this way, are *relative* rather than *absolute* measures.

4.6 Marcus's algorithm

If there are n different stimuli, they must all be presented to a subject as both A and B in the sequences, so that there are $n \times n$ trials. The final A - B intervals for each trial are noted in a $n \times n$ matrix of departures from isochrony, ignoring the diagonal (where each stimulus has been paired with itself). The matrix is termed [t]. The measured interval of the *ith* stimulus against the *jth* is t_{ij}. The hypothesis is that a stimulus has a unique P-centre, located at time $p_i + k$ from the onset of the *ith* stimulus. The k is a constant applying to all the experimental stimuli. Therefore the stimuli i and j will be perceived as isochronous when the P-centres are regularly spaced. The onset synchrony can be represented as $(p_i + k) - (p_j + k) = p_i - p_j$. The intervals that are measured between the onsets will result from the difference in the P-centres of the stimuli $p_i - p_j$. Additional variance will result from possible ordering effects (turning the knob further than actually desired) and random error. These can be considered to be independent of i and j. They also confound each other, and can be

represented by a single error term e_{ij}, with an average of K and a variance which is independent of the stimuli i and j. Thus

$$t_{ij} = p_i - p_j + e_{ij} \qquad 1$$

In the experiment the values of both t_{ij} and t_{ji} are measured. Based on this, estimates can be made of K and s_e for each run. The participants are told to aim for a K equal to zero. K should be normally distributed around zero. The value s_e^2 is a measure of the participant's variance in making the settings and has $w = (N - 1)/2 - 1$ degrees of freedom.

Let $$D_{ij} = (t_{ij} - t_{ji})/2 \qquad 2$$

and $$R^2 = (D_{ij} - p_i + p_j)^2 \qquad 3$$

A least squares P-centre fit is the set of p_i which minimizes R^2. These are computed by solving a set of simultaneous differential equations.
These N equations determine p_i, except for an arbitrary constant, chosen such that the sum of p_i = *zero.* The minimized value of R^2 is the total square error between the data t_{ij} and the P-centre fit p_i. The residual variance can be used to determine the subject's own replication accuracy.

As mentioned above, the P-centre calculated for a stimulus in this way is a relative, not an absolute measurement; and this makes it difficult to compare the P-centre locations for sounds across different experiments.

A way to improve the comparability of P centres across experiments is the use of a reference stimulus in all experiments. The reference noise used in this

thesis is 50ms of white noise, with 10ms linear ramping at onset and offset to avoid transient clicks and thuds.

The dynamic rhythm setting method was therefore used in this thesis for determining the P-centres of stimuli. There are two assumptions underlying this paradigm:

> *i) that participants can actually perform the dynamic rhythm setting,*
> *ii) that the rhythms subjects set reflect the P-centres of the signals involved, and not other attributes of the acoustic signal, such as the perceptual onsets, or some offset effects.*

The next two sections will cover both these topics. The first section describes Experiment 1, in which naive subjects performed a simplified dynamic rhythm setting task (simplified in that only one stimulus was aligned, instead of two). This was to test assumption (i), that the task is possible for the 'normal' population. In addition some variables which can be altered in a dynamic rhythm setting task were manipulated (for example tempo) to examine the effect on subject performance, and thus guide later experimental protocol. The second section addresses assumption (ii), that in a dynamic rhythm setting task it is the P-centres of the stimuli that subjects align. This is considered by examining relevant experiments in the literature.

4.7 Experiment 1 - aspects of a dynamic rhythm setting task

In this experiment[1] subjects performed the dynamic rhythm setting task, setting the reference sound to a perceptually even rhythm. If the subjects could perform this task accurately, then the intervals they set should be physically isochronous as well as perceptually isochronous, since only one stimulus is being presented. This first experiment has several aims, all connected with the first assumption of dynamic rhythm setting tasks.

1) Whether subjects can perform the dynamic rhythm setting task itself. If it is too difficult for subjects, it is not a suitable paradigm.

2) Associated with (1), the question of whether there are inter subject differences was considered. If subjects do vary significantly, then this represents a confounding variable which may affect the results.

3) Fraisse (1982) stated that not all rhythm tempos are equally easy for subjects to tap accurately to. At very fast and very slow tempos, subjects become inaccurate. A suitable tempo for the rhythm setting tasks was considered by using two different tempos.

4) Previous dynamic rhythm setting experiments have included in the subjects instructions some explanation that the subjects may clap, or tap along to help set the rhythm. The decision to do this is a pragmatic one, since subjects will tend to do so anyway (Seton 1989). The issue of whether subjects' rhythmic motor

[1] Lorna Atkins, Eddie Hamilton, Melanie Clutterbuck, Hayley Crawford and Sian Rees assisted greatly with the running of this experiment.

actions affected the accuracy of their settings, and thus whether subject movement represents an uncontrolled variable in the experiments was therefore examined. If there was a significant effect of tapping then this could be controlled for by explicitly instructing the subjects to tap (as Seton did) or by expressly instructing them not to. If there is no effect of movement, then subjects will be allowed to make gestures or not, depending on preference.

5) The changes in performance over time will be considered. This will be to provide a measure of learning over the trials. If subjects need a discrete number of trials before they acheive accuracy, then this must be observed in all experiments.

4.7.1 Method

4.7.1.1 Subjects

2nd year undergraduates (n=20) participated in the experiment, as part of a second year lab class. None were paid for their participation.

4.7.1.2 Apparatus

An IBM compatible 386 DX 33MHz PC, equipped with a 12 bit A/D, digital IO Data Translation DT2811 D to A board and a Fostex loudspeaker was used to present the sequences of stimuli to the subjects. Subjects started and ended the trials by pressing a programmable keyboard. The subjects adjusted the intervals between the stimuli by turning a potentiometer knob. The potentiometer knob could be adjusted by 100ms in either direction (200ms

overall). The smallest adjustment that could be made was 1ms. Whenever the knob was not moved for longer than 50ms the interval was recorded on the PC. The final interval was also recorded.

4.7.1.3 Stimulus

A 50ms burst of signal correlated noise with 10ms ramping at onset and offset was the experimental stimulus. This was the reference sound which was used in all thesis experiments.

4.7.1.4 Design

In a dynamic rhythm setting task, one sound (A) is always presented at regular intervals. This is the A - A interval. The second sound (B) is introduced at some point in the A - A interval. The position of B relative to the preceding A stimulus is represented as A - B. The initial A - B interval, and thus the initial B - A interval, is randomized. The subject adjusts this A - B interval until they are satisfied that the rhythm is perceptually regular. Since the actual A and B stimuli in this experiment are identical, subject should set the A - B interval to half the A - A interval. The rhythm should sound most regular when the stimuli are physically evenly timed.

There were four factors varied experimentally in this experiment - the tempo of the rhythm, whether or not the subjects were told to tap, the performance over the five trials, and the subjects themselves. The factor of **tempo** was varied by having two A - A intervals. In one condition the subjects set the rhythm at A - A = 1200 ms. In the second condition, A - A = 800 ms. The factor of **tapping** was varied by requiring half the subjects to clap/tap/stamp along to help them set the rhythm; the other half were explicitly told to make no actions at all. The

factor of **trials** had five levels (each subject completing five trials). The factor of **subject** was varied using random blocks.

The design was thus mixed random blocks and repeated measures, 2X2X5 factorial.

4.7.1.5 Procedure

The subject was seated in a sound treated room, and instructed that they were going to hear a sound like a drum beat from the loudspeaker. They were instructed to turn the knob, which would vary the interval between successive noises, until they felt that the rhythm of the sounds was **regular** in timing and evenly spaced. The aim of the experiment was for the subjects to set perceptually isochronous rhythms; since this term is not meaningful to most subjects the subjects were told that the rhythm to aim for was that of walking. This was mentioned explicitly to the subjects, along with the suggestion that they should alter the interval between the sounds until they matched the timing of "left...right...left....right" as called to a marching army. They were instructed to be as accurate as possible, and when they were satisfied that the rhythm was even, to press the programmable keyboard to stop the trial.

Depending on whether they were assigned to the 'tapping' or the 'no tapping' group, they were also instructed to make motor gestures to help them set the rhythm; or to make no movements at all.

4.7.2 Results

One subjects' results were discarded because this subject could not reach a satisfactory even rhythm in any trial. The table below shows the means of the

deviations from isochrony of the intervals set by the subjects. This is calculated by subtracting the (A - A)*0.5 value (perfect isochrony) from the interval set by the each subject, and transforming this into an absolute value. This provides a positive measure of the difference between the interval the subject sets, and isochrony. This value is thus a measure of how accurate the subjects were, the nearer this value being to zero the more accurate the subjects. This method controls for the subjects making settings that are both larger and smaller than perfect isochrony, which they will do if they are attempting to acheive perceptual regularity. If these raw intervals are analyzed, the mean will always be equal to isochrony, and any significant differences between groups will be lost. This method also enables the intervals set in the two tempo conditions to be directly compared. Table 4.1 below shows the means and standard deviations for the four groups.

	A - A = 800ms	A - A = 1200ms
Tapping	mean = 1.37ms S.D. = 53.40 S.E. = 5.34	mean = 3.30ms S.D. = 16.57 S.E = 1.657
Not Tapping	mean = -17.90ms S.D. = 56.50 S.E. = 5.650	mean = -0.13ms S.D. = 19.13 S.E. = 1.913

*Table 4.1 Means and standard deviations of difference from perfect isochrony (A-A)*0.5 of final settings made by all subjects over all trials, across all four conditions.*

4.7.2.1 Transforming the data

The variation in the size of the standard deviations shown in Table 4.1 indicated that the data needed to be transformed before parametric tests could be applied. This was done by taking the square root of the absolute deviations from isochrony. This reduced the variance (Table 4.2). This

transformation has reduced the variance, while differences between the groups are still apparent.

	A - A = 800ms	A - A = 1200ms
Tapping	mean = 4.233 S.D. = 3.299	mean = 3.296 S.D. = 1.493
Not Tapping	mean = 4.307 S.D. = 3.397	mean = 3.346 S.D. = 1.744

Table 4.2 Means and standard deviations of transformed (square root) absolute deviations set by subjects in four groups.

4.7.2.2 Subjects

The transformed deviations were regressed against the predictors **subjects** using multiple linear regression with subjects entered as dummy variables. The model fitted was

$y = c + n(ax)$

for the n=20 subjects. The dummy variables **subject** were not significant predictors of the transformed absolute deviations from isochrony set in the rhythm setting task. This indicates that the subjects were not making consistently different settings from each other. Subjects were not thus analyzed further.

4.7.2.3 Tapping, tempo and learning

A 2 X 2 X 5 repeated measures ANOVA was performed to determine the significance of the subjects' tapping or not, the tempo of the sequence, any change in their performance over the five trials, plus any interaction.

4.7.2.4 Main effects

The only main effect that is significant is **tempo** ($F_{1,100}$=5.14, MS = 26.94, p<0.05). **Trial** nearly reaches significance ($F_{4,100}$= 2.6, MS = 12.36, p=0.058). **Tapping** is not significant (F<1).

4.7.2.5 Interactions

The **tap X tempo** interaction is not significant (F<1).

Tap and trial

The **tap X trial** interaction is significant ($F_{4,100}$=3.89, MS = 20.38, p<0.05).

Figure 4.2 shows the means of the two tapping groups plotted against trial.

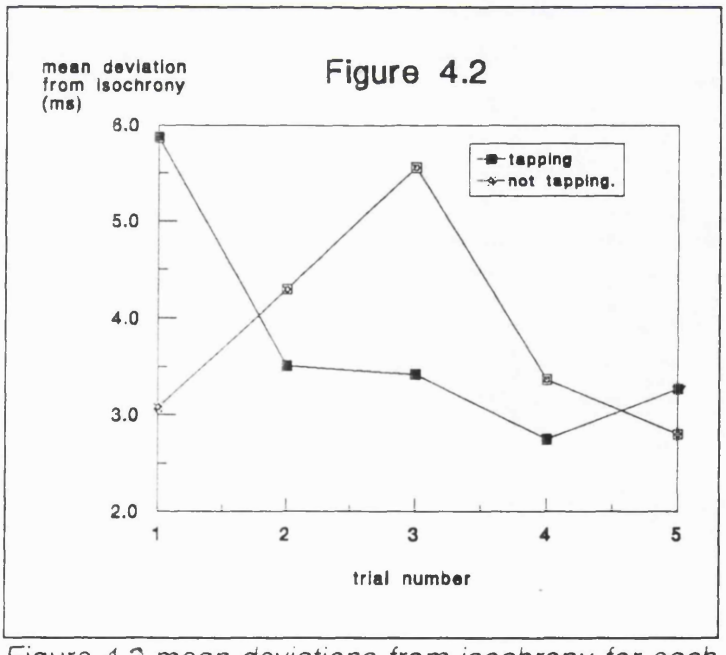


Figure 4.2 mean deviations from isochrony for each tapping group plotted against trial

The largest difference between the tapping and no tapping subjects is on the first trial in the sequence, when the subjects who were instructed to tap set intervals that vary further from isochrony than those told not to tap. The next large difference is on trial 3, where the subjects who are not tapping are making large deviations from isochrony.

Tempo and trial

The **tempo X trial** interaction is significant ($F_{4,100}$ = 2.86, MS = 14.99, $p<0.05$).

Figure 4.3 below shows the means of the two tempo groups plotted against trial. The main source of variation is the first three trials in the (A - A = 800ms) condition. The subjects are much less accurate on these trials than in the last two trials. The subjects in the (A - A = 1200ms) condition do not vary greatly over trials, and set lower deviations (are more accurate) than the subjects in the faster condition.

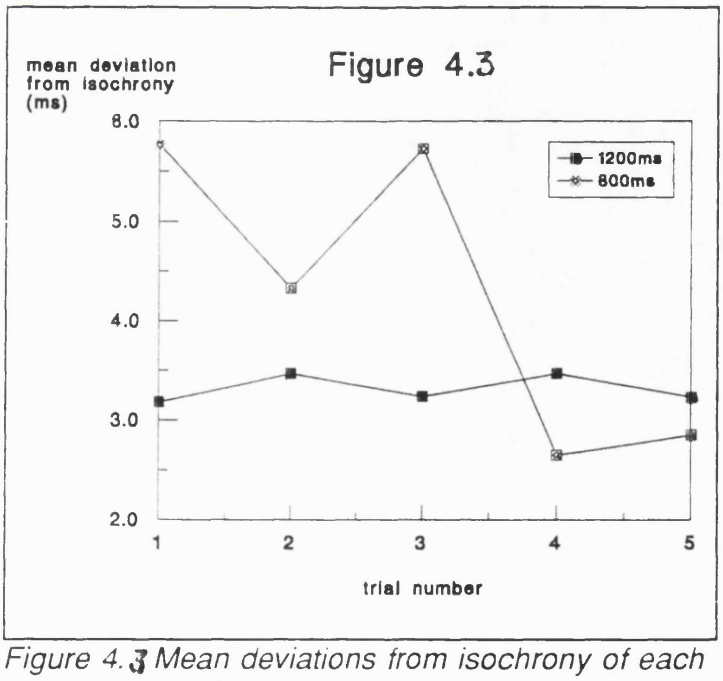


Figure 4.3 Mean deviations from isochrony of each tempo group, plotted against trial number

Tap and tempo and trial

This three-factor interaction is significant ($F_{4,100}$ = 4.19, MS = 21.95, $p<0.05$). The figure 4.4 below shows the mean deviations for each group plotted against the trials.

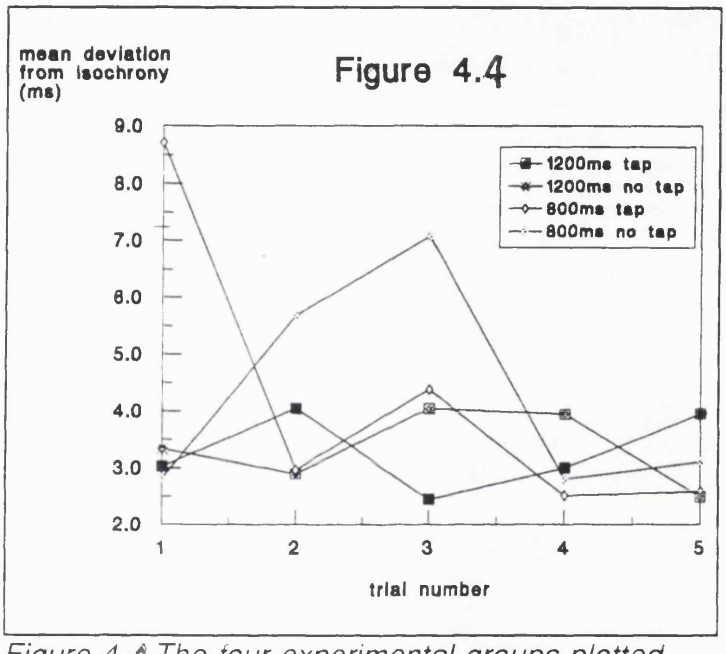


Figure 4.4 The four experimental groups plotted against trial number

On trial one, the tapping group (800ms) shows a large mean deviation from isochrony; the other three groups are similar in their mean deviations.

On trials two to five, the accuracy of the tapping (800ms) group improves, and is similar to both of the 1200ms groups, both tapping and not tapping.

On trials two and three the not tapping (800ms) group shows large mean deviations, indicating less accurate settings. Their deviations are more similar to the other three groups.

From this graph it can be concluded that at the 1200ms tempo, subjects make more accurate settings, whether or not they are tapping. At the 800ms tempo, subjects display large inaccuracies if they are not tapping; if they are tapping they start with large deviations, but their performance improves.

4.7.3 Discussion

These results show that the dynamic rhythm setting task can be performed by the vast majority of participants - subjects do not vary significantly in the settings they make. Only one subject had to be excluded from the experiment. This subject never acheived perceptual isochrony in the rhythm setting trials. The discovery of an individual who simply cannot perform the rhythm setting task to a level of satisfactory perceptual isochrony has a precedent; Seton (1989) reported one such subject.

Participants are slightly more accurate at a slower tempo. The errors in the A-A = 1200 ms condition were significantly smaller than those in the 800ms condition.

Tapping or not tapping was not significant as a main effect. The interaction of tapping tempo and trial indicates that:

> The production of rhythmic movements does not improve performance at the 1200ms tempo, nor does physical stillness degrade it.
>
> At the faster tempo tapping appears to improve performance over trials.

The factor of trial failed to reach significance as a main effect; This implied that overall the subjects showed no change over trials - that is, there was no main learning effect. The significance of the interaction of tap and tempo and trial is due to the tapping subjects at 800ms showing an improvement over time, and the non tapping subjects at 800ms showing degradation over time.

Thus the results of this experiment indicate that subjects can perform the dynamic rhythm setting task (with one noise) accurately (over the limited range of tempos tested); they are more accurate at a tempo of 1200ms, and tapping/no tapping generally has no effect on their performance; although it can improve performance accuracy at a faster tempo (800ms).

4.8 Implications for dynamic rhythm setting tasks

The dynamic rhythm setting task was thus adopted as an experimental method; in order to ensure that the subjects were performing accurately control conditions were incorporated into every experiment. While the P-centre algorithm does not use the trials where a stimulus is set against itself, these trials were always examined :

a) to familiarise the subject with the signals
b) to check that the subject was concentrating and making reasonable settings
c) to ensure that there was nothing about the signal which made it difficult to set to a rhythm.

Throughout this thesis, these trials where a sound is aligned against itself are indicated in the results tables by shaded cells. The mean intervals shown in these cells should always be near to the (A-A)*0.5 value, that is physical isochrony, if the subjects are correctly performing the experiment.

The next section considers whether the dynamic rhythm setting task facilitates the determination of the P-centre of a signal, or if the experimental protocol can affect the settings made. Do subjects sometimes set the stimuli to rhythms based on other attributes of the stimuli (such as the perceptual onsets, or the offsets, or the vowel onsets) depending on what they are instructed to do?

4.9 Do the experimental instructions affect the settings made in rhythm setting tasks?

This was investigated by Whalen Cooper and Fowler (1989). They were concerned by comments that subjects might be making rhythm alignment judgements that were influenced by the other qualities of the stimuli, rather than the P-centres. The explicitly suggestion was that subjects were aligning the **offsets** of the stimuli instead not the P-centres. To test this they expressly asked subjects in a P-centre determining experiment to align either the P-centre, the onset, the vowel onset or the offset of speech sounds. They found that subjects could not perform the offset alignment condition at all, and that in the other conditions they were making settings which were not significantly different from P-centre settings for the same stimuli. In a second experiment they altered the vowel duration of the stimuli, to make alignment by offset easier, by providing more distinguishable duration cues. They found that this was still not a possible task to perform, while the onset and vowel onset settings were still comparable to the P-centre settings for the stimuli.

From this they concluded that P-centres are the only perceptual events at which syllables can be aligned to sound regular, and that this meant that the P-centre settings were not affected by the subjects making other alignments, or by the experimental instructions.

Regrettably, given the questions that this experiment could have posed for the whole experimental paradigm of determining P-centres using dynamic rhythm setting tasks, they used only one naive subject. Three of the four participants were the authors. This is acceptable if the experiment investigates effects on responses which are driven by the stimuli. It is not acceptable in situations where the experimental variable measures whether subjects can or cannot perform a task. In addition, their initial A-B interval was always the same rather than random, leading to the possibility that subjects responses became influenced by a knowledge of how far the knob should be turned (that is, become stereotypic).

These results appear to be in contradiction to those of Seton (1989) who found that subjects could attend to different aspects of a signal in a rhythm setting task, depending on the instructions. In his case he used non-speech sounds, comprising a noise/periodic excitation compound (the periodic excitation was added into 250ms of white noise). The position of the periodic excitation within the white noise was varied. It occurred at either 0, 50, 100 150 or 250ms delay from the noise onset. Dynamic rhythm setting experiments were run twice with all four stimuli, with four naive subjects. On the first occasion the subjects were asked to make their rhythmic judgments based on the noise portion of the stimuli; on the second set of trials they were asked to concentrate upon the periodic components. If the subjects could perform both tasks, then different results would be found in the two different instruction conditions. This was indeed the case. In the noise condition, there was no significant difference in the P-centres set for the four different stimuli. In the second condition the later the periodic component occurred in the noise, the later the P-centres set for the stimuli, by all the subjects. Unfortunately Seton does not give any results for what would be found if the subjects were not told to attend to either component specifically, but merely asked to set the sounds to a rhythm as is normally

specified. Conclusions as to which component normally contributes more to the P-centre setting cannot therefore be drawn.

Reasons why his result differs so from that of Whalen et al can be considered. His use of naive subjects suggests that his results are more reliable; in the Whalen et al study the one naive subject does make different settings from the three experimenters. This cannot explain all the difference; the naive subject in Whalen et al did not show any signs of being able to perform the different settings. Another reason is Seton's use of non-speech sounds; while these were 'speech-like', they may have made the task easier by providing a clearer acoustic marker than 'vowel onset'. The noise and the periodic sections are acoustically much easier to separate than the segments of speech sounds. These stimuli may have provided clues that are more salient to naive listeners, to whom the instruction to attend to vowel onset is not necessarily a meaningful statement. Natural vowel onsets also display longer rise times than the 5ms of Seton's periodic component, again making their detection in the acoustic waveform harder.

The question of whether subjects can attend to different aspects of the acoustic waveform, and use these to make rhythm setting judgments remains open. The general consistency of P-centre results in the literature (with a few exceptions) indicate, as Whalen Cooper and Fowler (1989) state, that the P-centre is the most perceptually salient event to use to make rhythmic judgments, and that in different experiment subjects are not using different acoustic cues. However Seton's findings show that subjects can use different aspects of the signal to make their settings if so instructed.

4.10 General method used in this thesis

The general method used in this thesis followed closely that outlined in Experiment 1. The instructions remained the same through-out. The subjects were required to 'set the sounds to the rhythm, so that the beats sound even, like instructions to a marching army'. No mention was made of onsets/offsets, vowel onsets etc., and the assumption was made that, unless told not to, the subjects would set the sound using the most perceptually obvious event ie. the beat or P-centre. Since P-centres can be defined as what are measured in rhythm setting tasks, the best way of ensuring that the results are not misleading in any way is to run the experiments whilst always being careful to use identical verbal protocols, and continuous controls to check subjects' accuracy (see above).

4.11 Notes on the general method

4.11.1 Method

The method described in Experiment 1 is the general method used in all experiments described in this thesis. In each experimental trial two sounds are set against each other until the subject is happy that they sound even. The two sounds may be identical, or be different in some way. One of the pair is repeated at a constant interval (A - A interval) and the second occurs in between. The interval between the first and second sound (A - B interval) can be varied by the subject using a potentiometer knob. The starting A - B interval is always randomly set so that the subject cannot simply get used to turning the knob by a certain amount to acheive perceptual isochrony. The subject indicates that they are satisfied by the rhythm by ending the trial. If they cannot

make any satisfactory setting they inform the experimenter. All of the experimental protocol is controlled by a personal computer (PC). The subject carries out the trial unobserved, so they are free to make whatever gestures, sounds etc., they feel best help them make an accurate setting.

In this way, all the stimuli in an experiment are set against each other, and a common reference sound, to enable P-centres which can be compared across experiments to be calculated. The trials where the same sound is set against itself are included as controls/training trials. Each combination of stimuli pairs are presented to a subject several times in a row in each block; this is to enable the subject to become accustomed to the stimuli, some of which could sound a little odd on first hearing. The actual number of trials varied across experiments between 5 and 10 trials per block. Time and the subjects' patience were the constraints.

In a couple of cases, namely Experiments Two and Four, a full P-centre determining paradigm was not employed. This was for two different reasons. In Experiment Two P-centres were not being calculated, instead the patterns of intervals were of interest. In Experiment Four the stimuli (infinitely peak clipped speech) were very unpleasant for the subjects to listen to for any length of time. In this latter case the stimuli were just set against the reference sound and the resulting pattern of deviations from isochrony analyzed.

This method of setting a stimulus against a common reference sound alone, and extrapolating from the resultant deviations from physical isochrony, is the method most often used by researchers in P-centres (Pompino-Marschall 1992, Cooper Whalen and Fowler 1986 1988, Fox and Lehiste 1987 - see Chapter 1 for more details). This is probably because it is a much faster method that does not produce unreasonable results. In an experiment will n stimuli, rather than $(n+1)^2$ experimental blocks, only (n+1)*2 sets of trials are needed.

However this method drastically reduces the degrees of freedom in the final P-centre calculation, and also means that an actual determination of P-centres as defined by Marcus cannot be performed. Generally Marcus and Seton have been the only ones to use the full method.

4.11.2 Subjects

The subjects used in these experiments vary. Only those who could complete the task in a pilot trial were used. Motivation and time were issues in the use of subjects; they had to be prepared to commit themselves to several hours of experimentation for the simplest experiment. For the initial experiments all the subjects were naive. Only subjects with 'normal' hearing (as tested with a Racal Amplivox Model 2150 meter) were run.

From Experiment Six onwards, the number of subjects was reduced from 3/4 to 2. This was to increase the ease of running the experiments. One of these two subjects was generally the experimenter. The reason for this was that in the initial experiments the experimenter knew quite explicitly what the experimental hypothesis meant in terms of the intervals set in the task (ie. that infinite peak clipping should affect the P-centre of a stimulus). It would thus have been unacceptable to use her as a subject. In the Experiments Six onwards, the experiments are calibrating the effects of manipulations upon the stimuli. The subject gets no feedback mid task on his or her performance, and thus has only their own settings to base their judgement of their performance on. The benefits of using the non-naive experimenter in these experiments were that she was skilled at the task, motivated to be accurate and could spare the time needed.

The use of fewer subjects meant that more experimental trials could be performed. Throughout the experiments tests were performed to check that the

subjects were consistent with each other, and individual differences such as those that Cooper et al (1986, 1988) reported were not influencing the results. If subject differences were shown, attempts were made to account for them.

4.11.3 Analysis

It must be noted that the experimental analysis employed varied from Marcus's in one important respect. Marcus mainly ran himself alone as a subject. For the reasons given above this is not an ideal but a pragmatic choice. In an experiment with n stimuli, leading to $(n+1)^2$ experimental conditions, he would have a number of trials in each condition. Instead of pooling together all the interval as set in each condition, and using these to calculate P-centres for the stimuli, Marcus would calculate P-centres using the raw intervals from the all first trials in each condition, then a P-centre from the second trial in each condition; he would thus calculate as many P-centres as he had trials in each condition. This way he would collect a range of P-centres for any one stimulus.

In the analysis used throughout this thesis, the raw intervals were first pooled for each condition. They were then tested for various attributes eg. differences between subjects, unreasonably large standard deviations, evidence that there was more than one rhythmic centre being employed as a cue by the subjects. If, for instance, there were subject differences a reason for these differences would be identified. If the results were so different as to suggest the subjects had performed in very different manners the results would be analyzed separately. The results were also analyzed for a significant effect of stimulus combination condition on the set intervals.

If all the evidence suggested that the intervals set were not varying between subjects, and were varying due to the stimuli, then the means were used in

Marcus's P-centre algorithm. Thus for each stimulus in each experiment a separate P-centre is calculated.

4.11.4 Equipment

Experimental stimuli were created on a Masscomp RTU 5.0 Unix system running the SFS speech editing/analysis software. This enabled the digitising of real speech and the synthesis of speech and non speech sounds, as well as the editing and manipulating of stored speech files to create stimuli that varied in amplitude profile, duration etc. The SFS library facilitated the writing of programs in "C" to extend the range of manipulations or other functions if none suitable existed, eg. the code for infinite peak clipping (Experiment 4).

The dynamic rhythm task was controlled by computer. A program was written to run under DOS on a PC equipped with SFS and a D to A board. Stimuli were transferred to the PC from the Masscomp via the ethernet network, and stored on the hard disk of the PC. The D to A board limited the sampling frequency of the stimuli to 20KHz; any higher rate was not converted veridically to an analogue signal.

In the experimental program (code in APPENDIX ONE), an SFS file was selected and two items in this file chosen. The A - A interval (see Method, this Chapter, for details) was input (which could not be shorter that the combined durations of the two stimuli). The initial A - B interval was selected randomly by the program. The trial was then run, with the subject indicating when they were satisfied with the intervals they set by hitting a programmable keyboard and stopping the trial. The program wrote the speech file name, the two items used in the trial, the A - A interval, the A - B interval, the number of overall stimuli repetitions (that is, how long it took to complete the trial), and the final A - B interval that the subject set to an output file.

During the trials the signal output from the D to A board was amplified by a QUAD 520 Power amplifier, and passed into a sound treated chamber. The stimuli were presented to the subject via a Fostex loudspeaker for Experiments One Two and Three, and using an ER-2 insert earphone for Experiments Four to Eight.

The knob the subjects turned to alter the A - B intervals was a potentiometer knob built in house (to these specifications), which could be turned 100ms in either direction (5V = 100ms). The smallest adjustment that a subject could make was 1ms. The subjects were instructed to re-centre before every trial. An LED display controlled by the computer instructed the subject when to reset the knob, when to start the trials etc..

In the individual experimental chapters that follow, the above method will be observed, unless otherwise stated. The precise detail of the method will not be gone over each time. Instead the Method section will give details of the stimuli construction and collection, the number of experimental trials, whether any trials were omitted, and the subjects who participated.

4.12 Summary

The problems associated with determining where in time listeners hear a sound occur were considered, and an indirect measure was proposed. This indirect measure of perceptual centres is the dynamic rhythm setting task. An experiment was described which aimed to test some of the basic assumptions made about this task (that the task is possible, that motor actions do not affect performance), and it was found to be a task which most subjects could perform accurately with one experimental stimulus. Further assumptions about the task are that subjects behave in the same way in the experiments and do not attend to different attributes of the stimuli when making their settings. These issues

were considered in the light of other research; there is still some confusion on these points. A general method was outlined, which avoids some of these problems by always giving subjects an initial pilot trial, increasing motivation, collecting many data points, and continually testing for consistency between and within subjects' performances.

Chapter Five

Experiment 2:

The perception and production of rhythmic sequences

Abstract
A basic assumption of most models of P-centre location (for example Morton et al 1976, Fowler 1979) is that the pattern of intervals which lead to the perception of regularity in sequences are the same as the patterns produced when speakers utter isochronous sequences. This experiment tested this hypothesis using speech from eight different speakers. The produced intervals were compared to intervals set by subjects in a perception experiment. There was a good correspondence between the produced and perceived intervals which supports the hypothesis. There was considerable variation across the speakers in the relative sizes of the intervals between words which lead to perceptual isochrony; these differences were consistent in both the perception and production task.

5.1 Introduction

Most models of P-centre location apply the phenomenon to both speech production and perception. Experimental work supports this. The findings of Fowler and Tassinary (1981), Howell (1984) and Rapp (1971) have indicated systematic asynchronies in the production of perceptually even speech. The size of the asynchronies correlates reasonably well with certain aspects of the syllables such as initial consonant cluster duration. These attributes have in turn been shown to affect the P-centres in perception experiments. This is thus congruent with many models which tie speech perception together with speech production. The articulatory approach to P-centres **requires** that the relationship between perception and production hold; this constraint does not apply to more acoustic approaches (Howell 1984, 1988 a&b, Marcus 1981).

The existence of the relationship is an issue mainly in terms of the differences between theories. The interpretation of the relationship between production and

perception does vary. Action-perception theorists conceive the relationship as a direct one, with the production of gestures driving the perception of rhythmic structure. Information about the signal is not processed in this model of the relationship. The articulatory model is more clear about how the speech signal is not processed than how perception is actually achieved.

Other theorists who model P-centre location in terms of intensity/time distributions (Howell 1984) have taken a different approach. Howell hypothesised that speakers produce rhythmically timed speech according to their knowledge of the energy distribution of the signal they produce. This was based upon an experiment in which speakers were instructed to produce vowels of different durations; their produced timing varied with the vowel durations. The work of Fox and Lehiste (1987) on timing of vowels of different lengths supported this result; they concluded that the P-centres of signals are influenced by the whole syllable. Although they do not state this explicitly, this conclusion implies that they broadly agree with Howell's position on the timing of produced sequences. What is not made explicit in this approach is precisely how this is achieved by speakers. For example, is the speaker's knowledge calculated in real-time, or is it based on stored knowledge about the intensity profiles of the syllables they produce, or the acoustic consequences of vocal gestures they make.

It could be hypothesized that in terms of rhythmic production of speech a *local model* of P-centre location (as described in Chapter 1) provides a more intuitive basis for the P-centre phenomena than a *global model*. An account of P-centre locations which relates them to punctate acoustic/articulatory events would map easily onto patterns of speaker's timing behaviour. Speakers uttering perceptually rhythmic speech would align the events that lead to the percept of P-centres to produce rhythmic speech.

Disagreements about the nature of the production / perception relationship are thus centred around the different theoretical frameworks within which speech is modelled.

The theoretical implications of a lack of production / perception relationship would be great, especially for the direct, action-perception account of P-centres. There is one possible threat to the validity of the studies which have investigated this relationship. Many have utilized an external rhythmic source for the subjects to speak in time with, for example a metronome. This explicit external rhythm may have induced production artifacts as the subjects attempted to align their utterances with the source. The literature on voice and action synchronization and the attendant physical asynchronies (for example Auxiette and Gerard 1992, Prinz 1992, Semjen, Schulze and Vorberg 1992 - see Chapter 4 for further details) indicates that this method can lead to problems, such as anticipation effects. A more significant problem is that yoked experiments in which the experimentally investigated produced speech is used in subsequent perception experiments are rare. Yoked experiments would enable direct comparison between production and perception.

Fowler (1979) did perform a yoked experiment. She analyzed the interval produced by a speaker for the typical anisochronies found when the words in an even sequence have different P-centres. She then created sequences, using examples of the produced speech; the sequences were either physically evenly timed, or timed with the same pattern of intervals that the speaker produced. There were thus two types of sequences - one physically regular, the other naturally timed. Subjects perceived the naturally timed sequences as more regular.

As mentioned in Chapter 1, Fowler does not state whether the rhythmicity of the original speaker's produced sequences was not verified, for example by an

independent listener. This could introduce artifacts in the measured intervals; measured physical anisochronies may not necessarily leading to perceptual isochrony.

Another problem is that P-centres were not determined in the perception experiment; since a rhythm setting task was not used, the offsets from isochrony which lead to perceptual isochrony cannot be established. If a more sensitive test were made of the intervals that result in perceptual regularity, would the same perception and production relationship be found?

In this experiment the production / perception relationship was investigated further, using perceptually isochronous speech, which subjects produced without an external rhythmic aid. That is, they had to generate their own internal rhythm. The onset to onset patterns that they produced could be studied. In a parallel study, samples of each speakers' utterances were set by subjects to a rhythm. The perceptual onset to onset values could thus be collected. The same pattern of onset to onset intervals should be found in production and in perception if the perceptual centres of speech signals govern the perception and production of rhythm when the production of rhythm is internally generated.

The experiment is broken down into two sections. The first section will describe the collection and measurement of rhythmic speech from different speakers. The second section will describe a dynamic rhythm setting task, in which perceptual isochrony will be acheived with the speech from the different speakers.

5.2 Experiment 2a Production

5.2.1 Method

Sequences of rhythmic speech were collected from eight subjects, three male and five female. Subjects were instructed to take a breath and then to repeat the sequence "one..two..one..two..etc" until they need to take another breath. As the subjects were instructed in the perception experiments (see Chapter 4) they were told to time the words 'evenly', that is at intervals which feel regular. Their rate of speaking was to be whatever speed felt most comfortable. They were instructed to speak faster or slower if the experimenter felt they were 'rushing' their words, or speaking too slowly. Speakers who spoke too fast would produce speech which could not be analyzed in terms of onset to onset intervals. Speakers who spoke too slowly would not produce a rhythmic sequence at all. Several training trials were given with the experimenter present providing feedback. The feedback provided was restricted to help with speaking rate, guidance as to the rhythm to be acheived, and encouragement. These were recorded and the subject could hear the tape back if they wanted to. When both the subject and the experimenter were happy with the performance, the subject produced some more sequences. The tapes were then listened to by two raters (one the experimenter), who chose the perceptually regular sections of the utterances - that is, selecting out the offset sections of the sequences, where the speakers slowed down, and the sections at the onset of a sequence where speakers would normally take a couple of utterances before producing rhythmic speech. The sections were then digitized at 20kHz and the onsets annotated by hand from an oscillogram display which afforded magnification of sections of the display. Initial deviations of the time / amplitude plot were used to denote the physical onset. This measurement technique was repeated to improve consistency and accuracy.

5.2.2 Analysis

From the spoken sequences "one two one two", it was the one-two and two-one onset to onset intervals that were of interest. The relative sizes of these two intervals was a measure of the physical anisochronies that results in perceptual isochrony for each of the speakers. Therefore, for each speakers, these onset to onset intervals had to be measured and the means calculated.

5.2.3 Results

The results of one speaker (SKS) had to be discarded. This speaker failed to produce a perceptually isochronous sequence. Although they tended to be confident at the time of production, they would agree on a later hearing that the sequences were irregular. After several attempts their sequences were not analyzed any further, although their speech was used in a later rhythm setting experiment. Speech from seven speakers (three male, four female) was therefore analyzed. The means and S.D.'s for these speakers are shown in Table 5.1 below.

	Interval Types	
Speaker	One - Two intervals (ms)	Two - One intervals (ms)
SHS	703.10 (45.30)	700.57 (30.32)
PH	722.15 (64.9)	630.57 (35.55)
LCE	681.85 (23.50)	682.38 (31.58)
SM	504.10 (41.70)	587.40 (28.44)
DG	613.24 (41.99)	660.04 (30.59)
WC	659.34 (36.23)	702.23 (39.93)
SR	742.66 (35.06)	800.53 (19.68)

Table 5.1 of means and S.D's of produced (one - two) and (two - one) intervals from seven different speakers

Table 5.1 shows that the speakers are producing perceptually even sequences with inter-onset intervals that vary across the two interval types and the speakers. Speakers LCE and SHS show no clear difference between the (one-two) and (two-one) mean intervals. Speakers DG WC SM and SR show larger (two-one) mean intervals, than (one-two). Speakers PH shows larger (one-two) mean intervals. The significance of these differences was tested statistically, using independent t-tests. Each speaker was examined separately, since all the speakers spoke at different tempos, and therefore direct comparison across speakers was not possible.

These produced intervals are plotted against speaker for each word type in Figure 5.1. Independent t-tests were applied to each set of intervals, to test the hypothesis that for each subject, the (one-two) and (two-one) intervals were equal in size. The results are shown in Table 5.2.

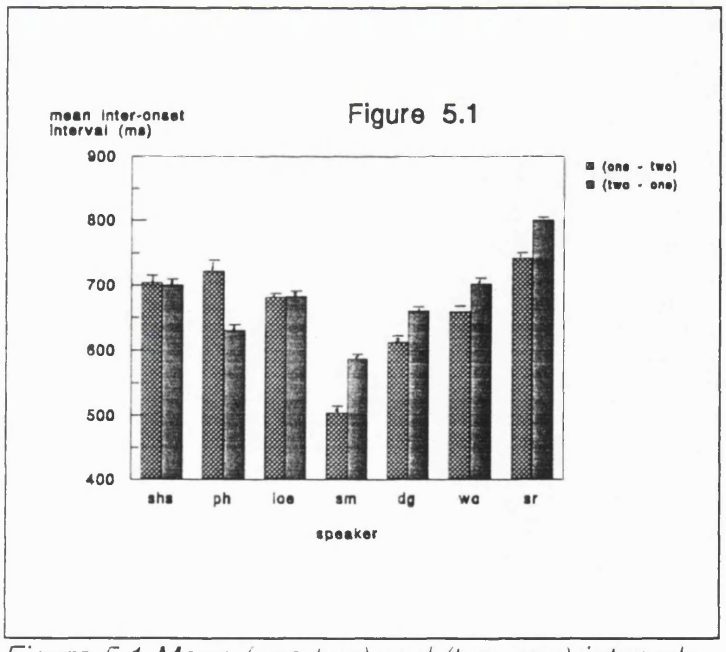


Figure 5.1 Mean (one-two) and (two-one) intervals, for each speaker in production

Speaker	T value	p value	degrees of freedom
SHS	t = 0.17	p = 0.86	df = 22
PH	t = -4.56	p = 0.0002	df = 20
LCE	t = 0.05	p = 0.96	df = 22
SM	t = -6.53	p = 0.0000	df = 26
DG	t = -3.93	p = 0.0004	df = 32
WC	t = -3.33	p = 0.0022	df = 32
SR	t = -5.76	p = 0.0000	df = 23

Table 5.2 independent t-test results comparing the mean (one-two) interval with the mean (two-one) intervals for each subject. Cases where the difference between interval types are significant are shaded.

Thus for the subjects PH SM SR WC and DG the differences between the (one - two) and (two - one) intervals are significant. As mentioned earlier, these differences are not all of the same magnitude or direction. While LCE and SHS show no difference between their intervals, subjects PH and SM show large, and opposite effects; while SR WC and DG show similar, smaller variation, all in the same direction.

The absolute sizes of the intervals are not considered further, since each speaker was producing sequences at different rates. Rather, this pattern of interval differences will be compared to those set in the perception part of the experiment. It is sufficient here to note that:

a) There are differences in several cases in the relative sizes of the (one-two) and (two-one) intervals.

b) The patterns of these differences varies across the subjects.

5.3 Experiment 2b Perception

In this part of the experiment, examples of each speakers speech was used in a dynamic rhythm setting task to determine whether the same patterns of interval types observed in production were replicated when the speech items were presented to subjects in a perception experiment. That is, are the perception intervals predicted by the production intervals.

5.3.1 Method

Examples of each speaker's (one)s and (two)s were selected from the digitized sequences used in Part a. This selection was only constrained by the speech items being those that were in the sequences used to determine the production intervals. No systematic criteria were possible with such as collection of natural, varying speech. One example of each speech item was used. Figure 5.2 shows the oscillograms of each item, for each speaker. Each was amplified to a maximum of 10V peak to peak. This was to control for the amplitude of the stimuli presentations. The stimuli were put into an SFS file, and transferred over to the audio lab, along with a reference sound (50ms of signal correlated noise, with a 10ms ramped onset and offset).

There were thus nine stimuli, a (one) and (two) from each speaker, and a reference noise. These were used in a rhythm setting experiment. For *each* speaker, their (one) and (two) tokens were compared to *each other*. This was to see if, in a rhythm setting experiment, with a standardized rate of presentation, the same pattern of (one) - (two) intervals as was found in the production part of this experiment, would be evident, implying differing P centres. Three subjects performed the experiment, with six trials in each experimental block.

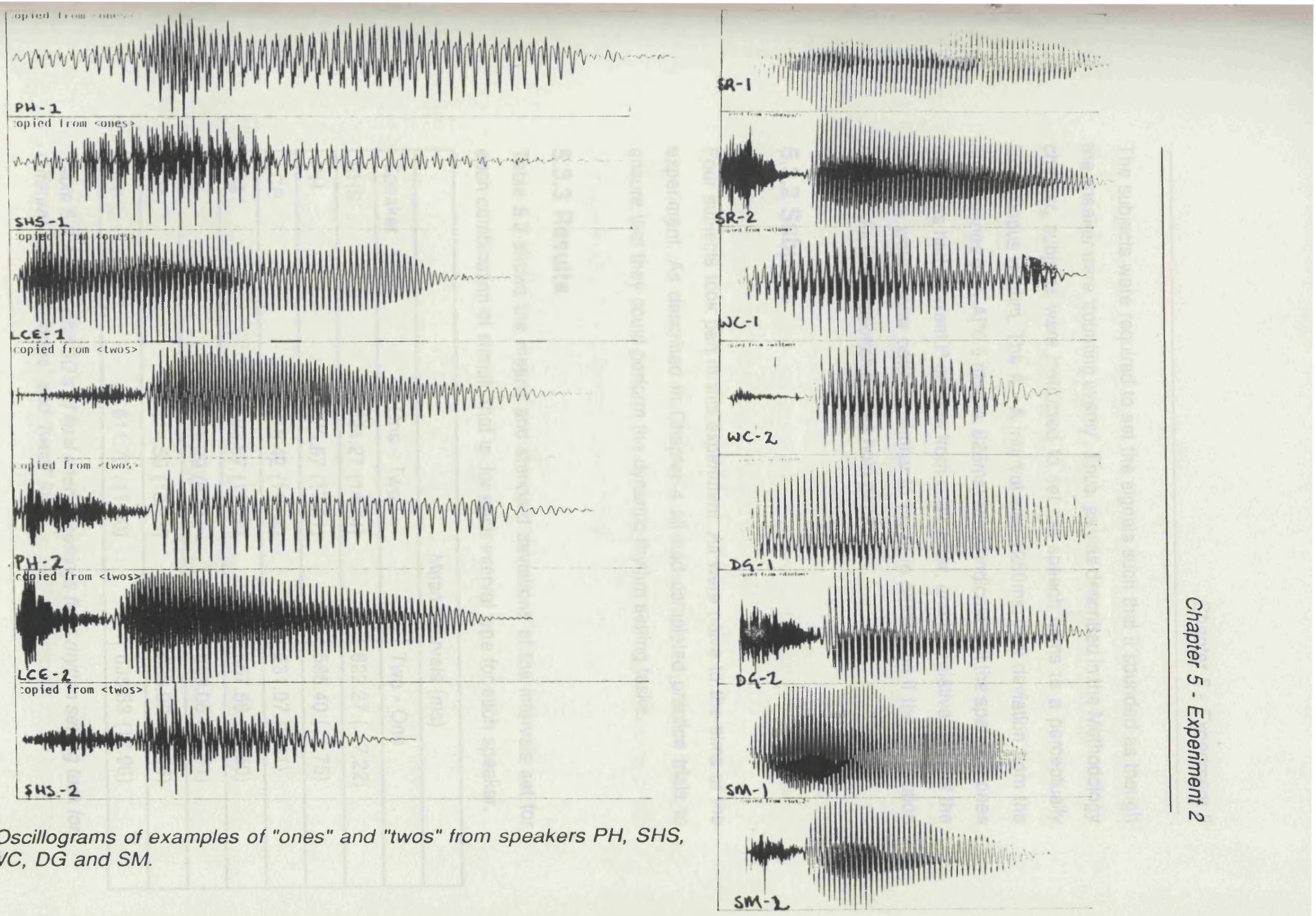


Figure 5.2 Oscillograms of examples of "ones" and "twos" from speakers PH, SHS, LCE, SR, WC, DG and SM.

The subjects were required to set the signals such that it sounded as though the speaker were 'counting evenly'. Thus, as was described in the Methodology chapter, subjects were instructed to set the speech items to a perceptually isochronous rhythm. The A - A interval was 1250ms. Any deviation from the final settings of [A-A]*0.5 (that is, 625ms) would indicate that the speakers ones and twos had different P-centres from each other, and the relative sizes of the intervals would show patterns similar to those in perception if the perception and production relationship does hold generally.

5.3.2 Subjects

Four subjects took part in this experiment. All were naive to the aims of the experiment. As described in Chapter 4 all had completed practice trials to ensure that they could perform the dynamic rhythm setting tasks.

5.3.3 Results

Table 5.3 shows the means and standard deviations of the intervals set for each combination of stimuli, that is, for each interval type for each speaker.

	Mean intervals (ms)	
Speaker	One - Two	Two - One
SHS	624.27 (14.74)	622.27 (17.22)
PH	675.67 (17.21)	585.40 (11.75)
LCE	615.42 (11.04)	631.07 (8.61)
SM	593.67 (17.74)	657.56 (19.50)
DG	612.29 (8.24)	639.06 (12.91)
WC	614.59 (13.13)	633.00 (15.66)
SR	611.76 (11.00)	635.83 (17.96)

Table 5.3 Means and SD's of final interval settings from rhythm setting task for different speakers "ones" and "twos" (ms)

The pattern of different one-two and two-one intervals shown in the production data in Table 5.1 is shown again in Table 5.3 above. For example, the intervals are different for the speech of speakers PH and SM, and the variation of interval size is in opposite directions.

The speech of PH, which exhibited the largest differences in interval size in the production data, is here set by subjects to have the largest deviations from absolute isochrony (625ms).

The speech of SM, as in the production data, shows the next largest deviation from isochrony. The (two - one) interval is still the larger.

The speech of LCE shows slightly different (one - two) and (two - one) intervals when set to a rhythm by subjects; this precise pattern was not found in the production data. The difference is very small.

Subject SHS shows, as in the production data, no difference between (one - two) and (two - one) intervals.

The speech of subjects DG WC and SR all showed smaller (one -two) intervals and larger (two - one) intervals in the production sequences; this pattern of intervals is maintained in the settings made for all three subjects in the perception task.

Thus the within subjects examination of the intervals shows a similar pattern of results in the production data. Figure 5.3 shows these perception intervals graphically.

The statistical significance of these differences in interval size was investigated in two ways. The first test was to establish that across the speakers, there were

differences in the (one - two) and (two - one) intervals set by the subjects - that is that perceptual isochrony was not based upon physical isochrony (as the perception results implied). The second test was to confirm the hypothesis that there are differences between the speakers in the patterns of intervals that result in perceptual isochrony (again based on the production data).

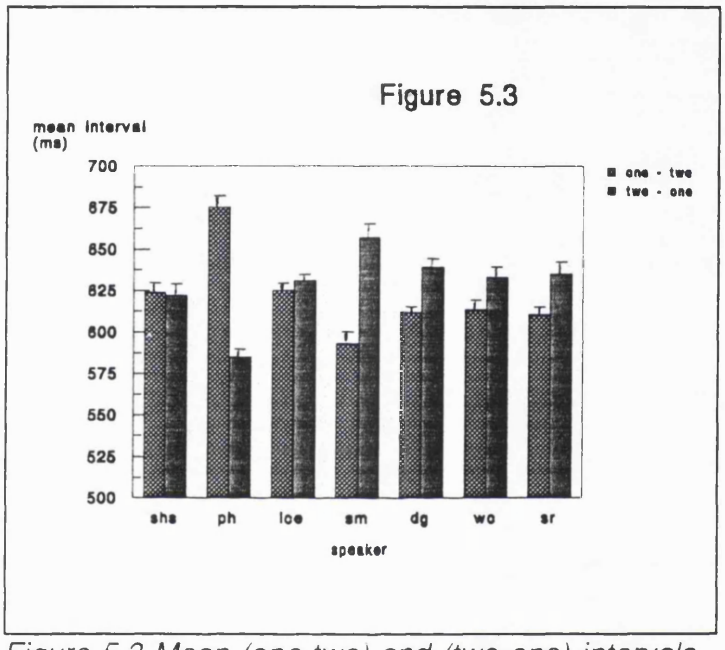


Figure 5.3 Mean (one-two) and (two-one) intervals for each speaker in perception experiment

5.3.2.1 Interval differences

The intervals set by the subjects for each "one" "two" stimulus were regressed against speaker, subjects, and the interval type; (one - two) or (two - one). This was performed with multiple linear regression, fitting the equation:

$$y = c + \alpha(x_1) + \beta(x_2)$$

where y=intervals set, x1 = speaker, x2 = interval type

The regression fit had the equation:

Intervals set = 617 - 0.733(speaker) + 7.95(interval type)

The only significant predictor was **interval**, ($t = 2.15$, $df = 1, 203$, $p < 0.05$). The predictor **speaker** was not significant ($p>0.05$). To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\gamma(x3)$ entered into the above model); this was not significant ($p>0.05$).

This regression shows that if the raw (one - two) and (two - one) intervals for each speaker are considered, then the only significant factor is the nature of the interval - that is the (one - two) and (two - one) intervals are different in size. This suggests that perceptual isochrony is not based upon physical isochrony. This corresponds to the results of the production experiment.

5.3.2.2 Speaker differences

The above multiple regression of raw intervals against interval type, speaker and subject showed no effect of speaker differences. As speaker differences were apparent in the production data, this issue was considered further.

The deviations from physical isochrony for each raw interval setting were calculated. Thus for each (one - two) and (two - one) interval, for each speaker, the size of the difference from [A - A]*0.5 (625ms) was found. This leads to values that range around zero (with zero representing no difference from physical isochrony). All the values were rendered absolute to give a value that expressed, for each speaker, the amount of physical deviation needed from physical isochrony to lead to perceptual isochrony. This measure was used to control for the fact that the physical shift could be in either direction. This absolute deviation gave a measure of the deviations set by subjects to acheive

perceptual regularity. A comparison of these values across the subjects would give an indication of whether the subjects differed significantly in the deviations they set across the speech items from the different speakers. The table 5.4 below give the absolute deviations of the intervals set with speech items from each speaker.

Table 5.4 shows, as the means of the interval show, that the speakers differ in the amount of deviation from physical isochrony needed in order to achieve perceptual isochrony. The statistical significance of this observation was tested by regressing all the absolute deviations against interval type, speaker and subject. A linear regression was used.

speaker	mean and SD of absolute deviations (ms)
SHS	12.733 (9.19)
PH	45.133 (15.53)
LCE	10.000 (7.16)
SM	34.600 (16.65)
DG	14.133 (9.60)
WC	13.900 (10.69)
SR	11.367 (9.52)

Table 5.4 mean and SD.s of absolute deviations from isochrony for (one - two) and (two - one) intervals for each speaker

The absolute deviations from isochrony were regressed against the predictors interval type, speaker and subject, using multiple regression, fitting the equation:

$$y = c + \alpha(x_1) + \beta(x_2)$$

where y=absolute deviation from isochrony, x1=speaker, x2=interval type

The regression equation was:

absolute deviation from isochrony = 31.4 - 2.6(speaker) - 2.52(interval type)

The predictor **speaker** was significant ($t = -4.56$, $df = 1,203$, $p < 0.05$). The predictor **interval type** was not significant ($p>0.05$). To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\gamma(x3)$ entered into the above model instead of x2 and x3); this was not significant ($p>0.05$). Subjects were thus consistent with each other in their settings.

The non significance of the predictor **interval type** implies that the subjects were performing the task correctly. If they were aiming to achieve isochrony, the (one - two) intervals they set would complement the (two -one) intervals set. Thus if all the subjects set the (two - one) intervals for one speaker as large, then they should set the corresponding (two - one) intervals small. The finding that the absolute changes made by all the subjects does not vary with interval type confirms this.

The significance of the predictor **speaker** means that the absolute deviations show in table 5.4 are varying significantly across the speakers. Thus large deviations are set for speakers PH and SM (45.133 and 34.600 ms respectively). Smaller deviations were set for the other subjects.

5.4 Comparison of perception and production results

To compare the perception and production data from each speaker, the mean (one - two) interval was divided by the mean (two - one) interval, to give a value which represented the ratio of (one - two):(two - one) intervals. If the

mean intervals were equal in size, then the ratio would have the value 1.0. If the (one - two) interval was larger than the (two - one), the ratio would be larger than 1.0; if the (two - one) interval was the largest, the ratio would be smaller than 1.0. These values were calculated for each speakers' perception and production data, and are shown in Table 5.5

	(one - two):(two - one) interval ratios	
	production	perception
SHS	1.042	1.003
PH	1.145	1.154
LCE	0.992	0.975
SM	0.858	0.903
DG	0.929	0.958
WC	0.939	0.971
SR	0.927	0.962

Table 5.5 ratios calculated from (one - two) and (two - one) intervals in both perception and production for all speakers.

The method of comparing the ratios of the mean intervals, rather than the raw intervals, was used for several reasons. It reduces the two intervals to a single value, so that perception and production can be compared. It is also a method for comparing the patterns of intervals across subjects that controls for the speakers producing speech at different speeds in the production experiment, and the experimentally constrained tempo in the perception experiment

If there is a relationship between perception and production that holds even when subjects are not producing metronome speech and when the perception measure is gathered in a sensitive rhythm setting experiment, then certain predictions can be made. The ratio of the mean (one - two) interval over the

mean (two - one) interval from the production experiment (the one-two:two-one ratio) should have a linear relationship with the one-two:two-one interval ratio set in the perception experiment.

The slope of the relationship is harder to predict; since the production ratios are based on many uttered intervals and the perception interval upon just one instance of each item repeated many times. For an ideal production / perception relationship the slope should be close to 1. It would be difficult to disambiguate deviations from this as either genuine production / perception differences or variation due to the difference numbers of speech items represented by each ratio. This being the case, the relationship should have a slope that is not too removed from 1; the slope should, of course, be positive.

This hypothesis was tested by plotting the perception ratios against the production ratios. This relationship is shown in Figure 5.4 with the regression line added.

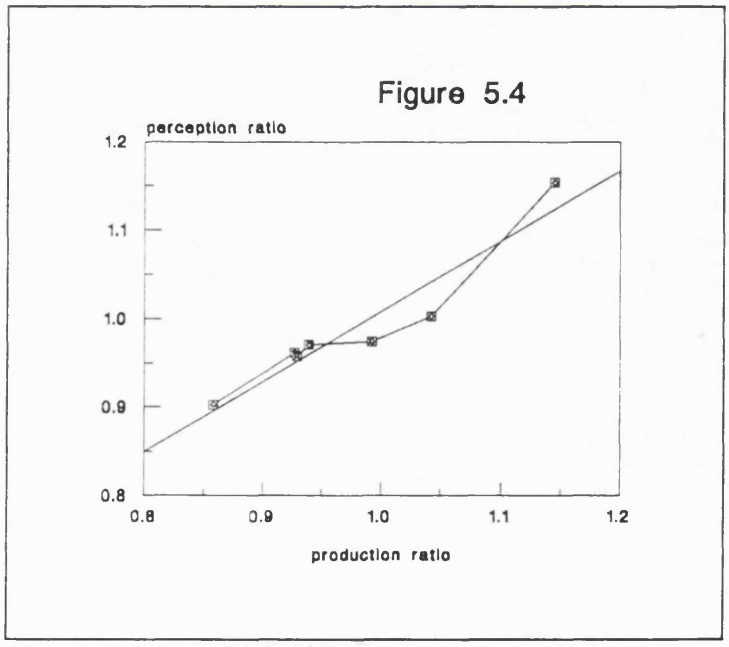


Figure 5.4 Plot of the ratio of perception intervals against the ratio of production intervals for each speaker

The relationship looks linear; the significant of this relationship was calculated by regressing the perception ratios against the production ratios. This give the slope of the relationship, as well as a measure of how much of the perception variance is accounted for by the production.

Thus a linear regression was performed of the perception ratios upon the predictor production ratios. The equation fitted to the data was:

$$y = c + \alpha x$$

where y=perception ratio and x=production ratio

The regression equation is
perception ratio = 0.219 + 0.789 production ratio

The predictor **production ratios** is significant ($t_{1,5} = 6.68$, $p < 0.05$). This predictor accounts for 87.6% of the **perception ratio** variation.

The slope is positive, and has a gradient of 0.789. This deviation from 1, as outlined in the section above, could be due to slight differences between produced and perceived rhythm; the deviation could also be due to the different numbers of speech items that are represented by each ratio.

The intercept *c* is not equal to zero; this is probably due to the tempo differences between the production and perception experiments. In production the limits of possible tempos, and thus interval sizes, was far greater than in the perception experiment. In fact the narrow tempo range of the perception experiment may have constrained the possible intervals set.

There is a significant relationship between the production and perception ratios. The ratio of interval sizes which a speaker produces is a predictor of the intervals which a listener will set in a rhythm setting experiment. The rhythmic

centres of a speakers utterances are encoded in the acoustic signal and can be used by a listener to set the sound to a rhythm in a rhythm setting task, at a fixed tempo.

5.5 Conclusions

The timing of naturally produced, perceptually isochronous speech corresponds closely to the timing set by subjects in perception results. The anisochronies observed when speakers produce even sounding speech are replicated when listeners have to set samples of these utterances to a perceptually even rhythm.

Further to this, the actual pattern of intervals varied across subjects - in both perception and production - for phonetically identical sequences. This implies that some attribute, or attributes, varies across subjects; this affects their produced timing; it also is expressed in the physical signal they utter, such that later perceptual tasks are affected. Implicit in these findings is the suggestion that the different speaker syllables have different P-centres.

In the next chapter, these differences will be addressed in more detail. The issue of whether any current models of P-centre location predict these results will be examined. Can Marcus's model account for these differences, or can a syllabic centre of gravity explanation predict these findings? Finally, a psychoacoustic model of perceived onset (Vos and Rasch 1981) will be considered.

Chapter Six
Experiment 3
Different speakers, different P-centres?

Abstract
The different interval patterns which were found to underlie rhythmic speech across speakers in Experiment 2 were investigated to establish whether this variation was caused by P-centre differences across speakers. The speech items "one" and "two" from four speakers (three were those used in Experiment 2) were used in two sets of dynamic rhythm setting tasks to determine P-centres for the items. These were compared, where possible, to the production intervals gathered in Experiment 2. P-centre differences were found to underlie the observed interval differences. Implementations of the P-centre models of Marcus (1981) and Howell (1988 a,b) and the perceptual onset time of Vos and Rasch (1981) were used to test each model's predictions against the P-centres for the speech sounds. Both Marcus's and Howell's models performed well, and the reasons for this are examined. P-centres were established for the speech items "one" and "two" from a further four speakers in reduced dynamic rhythm setting trials, and these P-centres compared to the production intervals from Experiment 2.

6.1 Introduction

Experiment 2 showed that there is a close relationship between perceived and produced isochronous rhythms across speech from different speakers. This is in accordance with previous studies which have indicated the strength of this relationship (Howell 1984, Fowler 1979, Fowler and Tassinary 1981, Fox and Lehiste 1987). The results of the Experiment 2 also suggested that different speakers produced different physical intervals to achieve perceptual isochrony. These intervals were reflected if the speech from each listener is used in a rhythm setting experiment. A conclusion that can be drawn from this is that the different speakers were producing the phonetically identical monosyllables with differing P-centres.

This conclusion was backed up by the casual observation that the oscillograms of the speech varied markedly across speakers in many acoustic features (duration, rise time). When listening to the different speakers the acoustic patterns of the speech were varying; some speakers producing "one" with a very 'soft' onset, others with an abrupt onset. Figure 5.2 in the previous chapter includes oscillograms of the "ones" from eight different speakers, and marked differences can be seen in their amplitude profiles and onset characteristics.

To investigate this, the P-centres of the some of the signals used in the previous rhythm setting experiment were determined experimentally. This involves setting all the sounds against each other, and thus for n stimuli there are (n x n) trials. The P-centres were also compared to the timings the speakers produced when uttering perceptually regular sequences (where the data was available). If P-centres are the events which are regularly timed in perceptually even sequences, then there should be a reasonable mapping between the two. If there is no clear relationship, this is a problem for the concepts behind the study of P-centres.

No current models of P-centre location were developed using speech from more than one speaker. The experimentally determined P-centres were therefore compared to predicted P-centres for these stimuli based on the models of Howell (1988 a&b), Marcus (1981), and the psychoacoustic non speech model of Vos and Rasch (1981). The most general model should give the best fit to the data.

Two separate experiments (Method I and Method II) are described here; the first is two full dynamic rhythm setting tasks using two types of speech item from four different speakers; the second is four smaller dynamic rhythm setting tasks using speech from four other speakers.

6.2 Method I

6.2.1 Stimuli

The speech items "one" and "two" (uttered by three speakers PH, SHS and LCE) which had been previously collected in the perception / production experiment (see previous chapter) were used. In addition speech from an another female (SKS) was included. These eight speech tokens formed the experimental stimuli. They were used in a P-centre determining rhythm setting experiment. These were all digitized at 20kHz, and 10V peak to peak (to control for peak amplitude). Figure 6.1 shows the oscillograms of these stimuli.

6.2.2 Design

This experiment was unlike the Experiments 1 and 2b described previously. In Experiment 1 only one stimulus was set to a rhythm (the reference sound). In Experiment 2a pairs of sounds were set to a rhythm; for every pair of stimuli there were two blocks of trials, the first where one sound (A) was repeated at a regular interval (the A - A interval), and the second sound (B) was adjusted relative to A until perceptual isochrony was achieved and the A - B interval noted. In the second block of trials the two stimuli were exchanged. Thus each member of the pair was presented as both the regular sound, and as the adjusted sound.

In a full dynamic rhythm setting experiment, the same principle of each sound being presented as the regular sound (A) and as the adjusted sound (B) was observed; however in this instance all possible combinations of the n stimuli in the experiment were presented as both (A) and (B). Thus instead of n*2 blocks of trials, there were n^2 in total.

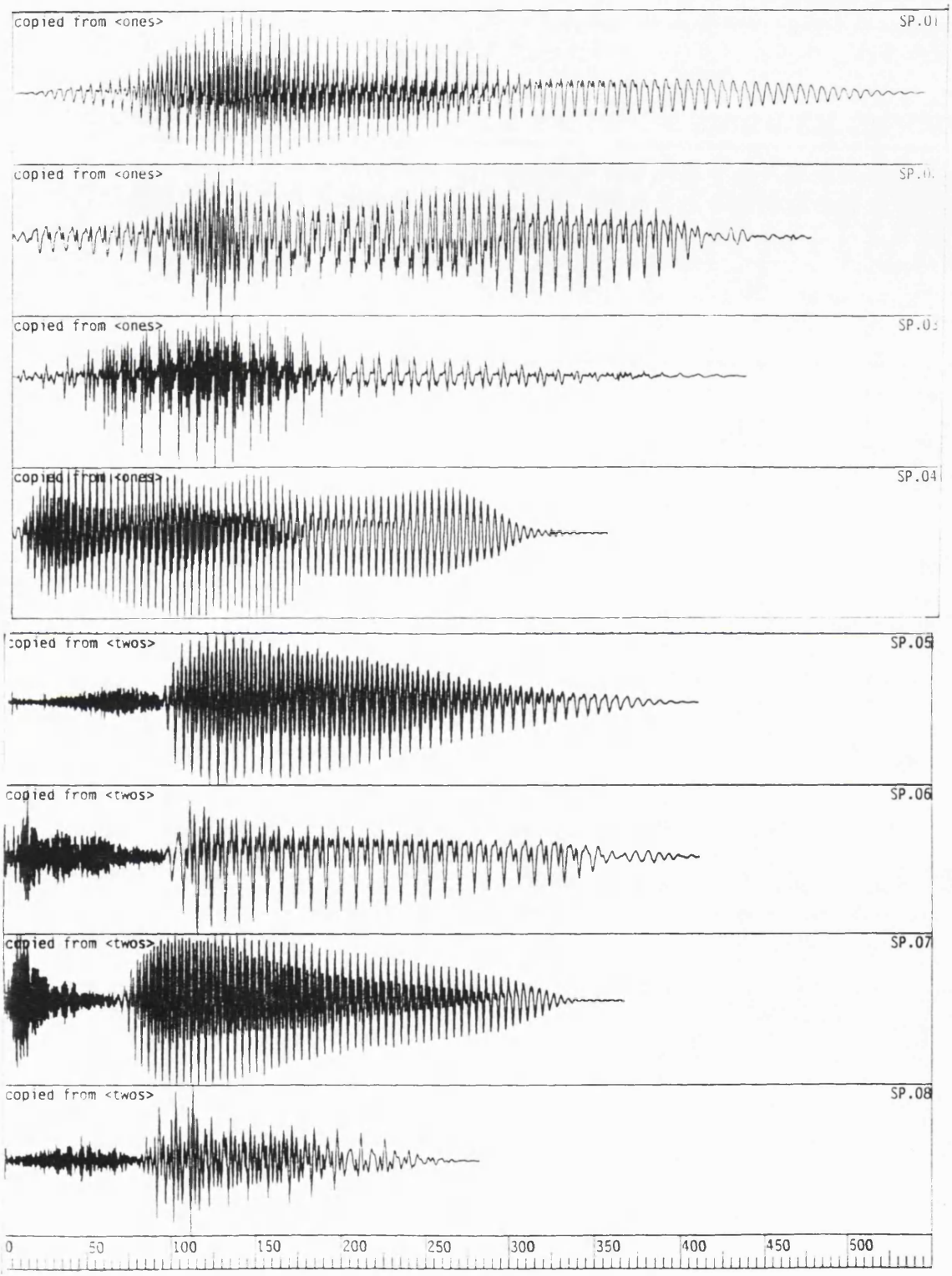


Figure 6.1 Oscillograms of examples of "ones" and "twos" from speakers SKS, PH, LCE and SHS.

This experimental method was therefore that which is described in the Methodology chapter (Chapter 4). The A-A interval was 1250ms. The initial A-B interval was randomized. The reference sound was used to enable comparisons to be drawn across experiments.

To reduce the number of trials, not all the stimuli were set against each other. This would have led to 81 sets of trials. Instead all the "one"s and the "two"s were grouped together as stimuli in two separate experimental sections, each including the reference sound. There were five stimuli in each section - leading to 25 experimental blocks per section. There were therefore 50 blocks of trials overall. Four subjects were run on the 50 trials (25 + 25), each subject completing 4 trials per block.

6.2.3 Subjects

There were four subjects, all those who took part in the perception section of Experiment 2. This was to enable the pooling of the results of the two experiments, and facilitate comparisons across the two experiments.

6.2.4 Analysis I

These results from Method I were analyzed in two ways. To demonstrate that the subjects were making consistent settings, and that there were significant differences in the settings made with different stimuli, a multiple linear regression was performed on the deviations from isochrony that the subjects set.

To calculate P-centres for the stimuli, the results were averaged across the subjects, and a matrix of mean final settings for the ones and the twos was completed. These were run through the Marcus P-centre algorithm - not to be

confused with his model - and the results compared to the experimental predictions that other models would make.

6.3 Results I

Tables 6.1 and 6.2 show the means and the standard deviations for the intervals set with all the stimuli combinations from the "ones" and the "twos"

stimuli "ones"	SKS	PH	SHS	LCE	ref
SKS	624.20 4.92	622.11 14.21	653.00 20.60	667.56 18.97	671.67 23.83
PH	640.44 21.31	627.78 6.51	669.44 25.81	693.67 26.14	703.11 24.35
SHS	581.00 9.21	568.00 13.83	622.89 5.99	648.44 9.42	659.44 15.91
LCE	584.78 16.05	567.89 23.88	601.89 11.53	624.89 10.13	646.22 13.22
ref	583.67 19.67	569.22 28.57	595.22 17.23	603.44 16.46	

Table 6.1 of means and SD's for all stimuli combinations for the different speakers "ones" (ms)

From the Tables 6.1 and 6.2 the shaded intervals, where the same sound is set against itself, are close to 625ms (physical isochrony). This shows that the subjects were successfully setting identical sounds (that therefore have identical P-centres) to physically isochronous rhythms. The standard deviations for these trials are small. The other intervals vary from 625ms in what appears to be a systematic manner, indicating shifts in P-centres across the stimuli. The standard deviations for these trials are larger, reflecting their difficulty.

stimuli "twos"	SKS	PH	LCE	SHS	ref
SKS	620.89 6.13	645.22 13.63	652.00 12.05	641.67 10.27	679.11 14.98
PH	615.89 7.49	627.00 4.90	643.56 7.92	631.78 11.94	658.89 20.06
LCE	599.22 14.18	617.33 7.26	625.11 2.42	607.67 5.72	645.33 12.35
SHS	612.00 15.31	618.22 17.73	642.11 9.73	620.22 3.42	661.33 16.53
ref	588.33 19.18	609.44 13.42	605.11 20.52	589.33 20.54	

Table 6.2 of means and SD's of final intervals settings for all stimuli combinations for speakers' different "twos"

6.3.1 Regression

The intervals set for both sets of stimuli were tested to ensure that the intervals set with the different stimuli combinations were statistically significant, and that the four subjects were not significantly varying in their settings.

To do this, the absolute deviations from isochrony were calculated for all the intervals. This was done by subtracting the [A - A]*0.5 value from every interval setting, and making the resulting value absolute. This gives a measure of how much a pair of stimuli had to be shifted from physical isochrony for perceptual isochrony to be achieved, regardless of the direction the shift was in. Tables 6.3 and 6.4 below show the means and the standard deviations of the absolute deviations from isochrony for the intervals set for the "one"s and the "two"s.

Tables 6.3 and 6.4 show that the mean absolute deviations from isochrony vary across the stimuli. The means and SD's for the pairing of identical stimuli (the shaded cells) are small, again indicating that the subjects made settings

approximating to physical isochrony when the P-centres of the stimuli were the same (as they should be with identical signals). The mean absolute deviations that are larger indicate that the subjects set these pairs of stimuli at a great offset from physical isochrony, indicating that the stimuli concerned have different P-centres. The standard deviations of these trials are larger.

"one"s	SKS	PH	SHS	LCE	ref
SKS	4.556 (1.236)	12.00 (7.081)	28.000 (20.688)	42.556 (18.974)	46.667 (23.828)
PH	21.889 (13.541)	5.889 (3.621)	45.556 (23.532)	68.667 (26.144)	78.111 (24.356)
SHS	44.222 (9.212)	57.000 (13.823)	4.778 (3.898)	23.444 (9.422)	34.444 (15.907)
LCE	40.222 (16.045)	57.111 (23.877)	23.111 (11.527)	8.556 (4.503)	21.222 (13.673)
ref	41.333 (19.672)	55.778 (28.569)	29.778 (17.232)	22.000 (15.788)	

Table 6.3 means and standard deviations of absolute deviations from isochrony of intervals set with "one"s stimuli (ms)

"two"s	SKS	PH	LCE	SHS	ref
SKS	6.111 (3.882)	20.444 (13.249)	27.000 (12.052)	16.667 (10.271)	54.111 (14.979)
PH	9.111 (7.491)	4.000 (3.24)	18.556 (7.923)	10.778 (7.965)	33.889 (20.059)
LCE	25.778 (14.175)	8.556 (6.044)	1.667 (1.658)	17.333 (5.723)	20.333 (12.349)
SHS	16.560 (10.77)	15.222 (10.269)	17.333 (9.274)	5.222 (2.587)	36.333 (16.53)
ref	36.667 (20.059)	17.111 10.084)	23.222 (16.107)	35.667 (20.537)	

Table 6.4 means and standard deviations of absolute deviations from isochrony of intervals set with "two"s stimuli (ms)

The absolute deviations from isochrony for both sets of stimuli were regressed against the stimuli combinations and the subjects to determine whether there was a significant effect of either factor on the settings that the subjects made. This was performed with multiple linear regression, and the equation fitted to the data was:

$$y = c + \alpha(x_1)$$

where y=absolute deviation from isochrony, x1=stimuli combination

6.3.2 "Ones"

The absolute deviations from isochrony for the "ones" stimuli were regressed in multiple linear regression with the predictors **stimuli combination** and **subject** (see above for equation fitted).

The regression equation was:

absolute deviation from isochrony = 16.7 + 1.32(stimuli combination)

The predictor **stimuli combination** was significant ($t_{1,211} = 5.55$, $p < 0.05$). To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was not significant ($p>0.05$). Subjects were thus consistent with each other in their settings. In the trials with the "ones" stimuli, there was a significant effect of the stimuli combinations upon the absolute deviations from isochrony in the intervals the subjects set.

6.3.2 "Twos"

The absolute deviations from isochrony set with the "two"s stimuli were regressed against the predictors **stimuli combination** and **subject** in a multiple linear regression (see above for equation fitted). The regression equation was:

absolute deviation from isochrony = 7.42 + 1.17(stimuli combination)

The predictor **stimuli combination** was significant ($t_{1,213} = 8.06$, $p < 0.05$). To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was not significant ($p>0.05$). Subjects were thus consistent with each other in their settings. In the trials with the "twos" stimuli, there was a significant effect of the stimuli combinations upon the absolute deviations from isochrony in the intervals the subjects set.

6.4 P-centre calculation

The mean intervals set by the subjects were therefore used in the P-centre algorithm, which returns a best fit for the P-centre for each stimulus, calculated from the matrix of intervals set (as shown in tables 6.1 and 6.2). The P-centres are shown in the table 6.5 below. A more negative value indicates a P-centre that is further away from the onset of the signal; a value nearer to zero indicates a P-centre nearer to the onset of the signal.

Since any one set of P-centres are all relative to one another, and in this thesis have all been calculated with respect to a common reference sound, the P-centres for each set of stimuli were adjusted such that the reference sound P-centre is always equal to zero. Thus the P-centres can be explicitly compared across sets (although they are still relative values, they are now relative to the same baseline).

Thus, the P-centres of the different speech items are varying across the subjects. The "one" with the latest P-centre is that of PH (-70.6ms); the earliest is that of LCE (-11.6ms). There is a difference of nearly 60ms between these two "ones". Variation is also apparent in the P-centres of the "twos". There is

a suggestion of a correlation between each speakers P-centres, except for speaker PH whose "one" and "two" P-centres vary by 40ms.

	P-centre (ms)	
Speaker	"ones"	"twos"
SKS	-55.2	-45.0
SHS	-26.6	-31.6
LCE	-11.6	-17.8
PH	-70.6	-30.6

Table 6.5 P-centres for "ones" and "twos" for four different speakers, as determined in dynamic rhythm task (ms) (relative to reference sound P-centre = zero)

Figure 6.2 shows the P-centres of the "ones" and "twos" for the different speakers; the nearer to zero the P-centre value, the nearer to the onset of the syllable the P-centre is.

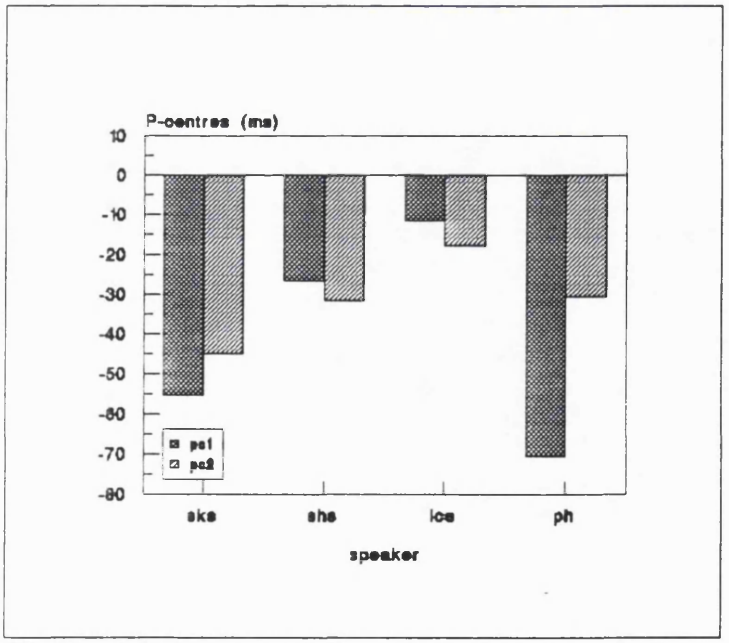


Figure 6.2 Measured P-centres for speech items "one" and "two" from different speakers

6.5 P-centres and production sequences

A direct comparison between the calculated P-centres for the "ones" and "twos", and the naturally produced timings reported in Experiment Two is not possible. This is despite three of the above speakers being those in the prior experiment. The comparison is difficult because of the time constraints which mean the "ones" and the "twos" were treated separately in this P-centre determining exercise. Thus though the P-centres are determined relative to the same baseline, the reference sound, the "one"s and "two"s P-centres were determined in separate experiments, and are based on the comparisons within speech types, not across them.

However the shifts in physical isochrony suggested by the P-centres can be compared to those observed in the production sequences, for the three speakers. The fourth speaker SKS produced sequences, that when later rechecked, were not perceptually isochronous. This does not alone suggest a production / perception association; rather it means that great care should be taken when collecting rhythmic speech from a speaker (see Chapter One and Experiment Two). In this case the subject did not utter rhythmic speech in the sense of speech that was very regular in any way. This sequence was collected without feedback from an observer. This regrettable fact highlights the importance of training and checking a speakers "rhythmic" speech.

For speaker PH, the Figure 6.3 below shows the pattern of mean (one -two) and (two - one) intervals as determined in the production of rhythmic sequences (determined in Experiment 2).

Thus the physical onset of the "one" is early with respect to the offset of the "two". This implies that relative to the "two" utterance, the "one" has a late P-

centre. Thus the physical onset of the "one" is started early, and the onset of the "two" brought forward, to achieve perceptual isochrony.

time ->

[one] [two] [one]

<-------722.15ms---><-------630.57ms----->

Figure 6.3 - mean onset to onset intervals for repeated "one-two" sequence spoken by PH.

The P-centres of "one" and "two" from PH bear out this prediction (although based upon just one example of each, calculated in different experimental blocks). The P-centre of the "one" is -70.6ms; that of the "two" 30.6ms (reference sound = 0ms). The "one" has a P-centre much later than the "two".

The production intervals for the speakers SHS and LCE, as described in Experiment 2 were not significantly different. Therefore the P-centres for the speech items from both speakers should be more similar than those of speakers PH. The P-centres of "one" and "two" for SHS are -55.0ms and -45.0ms respectively; those for speaker LCE are -11.6ms, -17.8ms. There is thus a difference in the P-centres of the "one"s and "two"s for each speaker, but these differences are small compared to speaker PH. The P-centres of the speech sounds thus reflect the pattern of onset to onset intervals speakers produce in rhythmic speech. P-centres are therefore representations of the basis of rhythmic centres in speech sequences.

6.6 Existing models of P-centre location

Can any of the existing P-centre models account for these findings? The P-centre predictions that each make were plotted against the observed P-centres, the regression line was calculated, and the amount of observed variance accounted for by each model compared.

First, versions of each of the models were implemented using software. The specific programs are in the APPENDIX.

6.7 Implementations of the models

6.7.1 Vos and Rasch

This model relates perceptual onset (defined in the same way as P-centres are in this thesis) to a threshold value that alters according to the peak intensity of the signal. When the signal crosses 15dB SPL below the maximum intensity value, perceptual onset occurs. This model was implemented with a program which converted the amplitude profile of a signal in Volts into dB re 1μV (using the 12-bit A-to-D values converted into volts, since 2047 bit values corresponds to 5V). The amplitude of the signal is expressed in volts in the SFS file; this is converted to dB using the equation:

$$dB = 20\log(V/V_{ref})$$

Where V_{ref} = 1μV

The peak intensity found, and the time, measured from the onset, that the signal passed -15dB re 1μV found.

6.7.2 Howell

This model relates the P-centre location to the distribution of amplitude over the whole signal duration. An increase in amplitude at the onset will shift the P-centre nearer to the onset; an increase at the offset will shift the P-centre nearer to the offset. There is no weighting function; amplitude changes at any point in the distribution have as much effect upon the P-centre of the signal. This model was implemented with a program that added up all the sample values of a signal from onset to offset, and used this value to calculate the point in the signal at which half of the total energy had passed. The implementation thus integrated all the sample values (expressed as 12-bit A-to-D values), calculated the total number, calculated half of this total, integrated the signal again until the half-total point was reached, then returned the point in time when this occurred. This can be expressed as an equation:

$$t_{centre\ of\ gravity} = t_{((\Sigma i\text{-}in)/2)}$$

Where t = time values and i = sample values.

This value would shift with corresponding changes in the amplitude over the entire signal as Howell's model predicts. This implementation was thus used to test whether the P-centres of stimuli varied with this representation of their amplitude distribution.

6.7.3 Marcus

Marcus's model relates P-centre location to the durations of two different aspects of a syllable - the onset and the rhyme. The model can be expressed as:

$$P\text{-}centre = \alpha x + \beta y + k$$

Where x and y represent initial consonant or consonant cluster duration, and vowel plus final consonant duration respectively. α and β are parameters of the

model which have the values 0.65 and 0.25 respectively. *k* is an arbitrary constant reflecting the relative nature of P-centre values.

The vowel onset is defined as the peak increment of mid-band spectral energy (between 500-1500Hz). The implementation took the signal - filtered between 500-1500Hz bandpass filter - and found the peak increment in amplitude (in Volts) within a time frame of 5ms. This was used to provide the *x* and *y* values in the equation above to predict a relative P-centre location for the syllable. The P-centre value can be correlated with the observed P-centres to determine its predictive power.

NB. The values of the all the model predictions are expressed as negative values. This is because they are conceived as measured in terms of distance from the onset of the signal, as are P-centre in this thesis. Thus a signal with a P-centre simultaneous with the onset would have a P-centre of 0ms. A signal with a P-centre occurring 50ms after the onset would have a P-centre of -50ms. This description will apply to the various P-centre model predictions, which will be expressed as negative values, the more negative, the further the predicted P-centre lies from the onset.

6.8 Regression of observed P-centres against model predictions

6.8.1 Vos and Rasch

Table 6.6 below shows the Vos and Rasch perceptual onset model predictions for the P-centres of the "one"s and "two"s stimuli.

Figure 6.4 shows the observed P-centres plotted against the Vos and Rasch model predictions.

Vos and Rasch Model (ms)	"One"	"Two"
speaker SKS	-59.3	-57.7
speaker SHS	-30.6	-0.5
speaker LCE	-5.4	-0.2
speaker PH	-15.9	-0.6

Table 6.6 Vos and Rasch threshold model of perceived onset predictions for the experimental stimuli (ms).

The two sets of data were regressed together in a linear regression fitting the equation:

$$y = c + \alpha x$$

where y=P-centre and x=Vos and Rasch model prediction

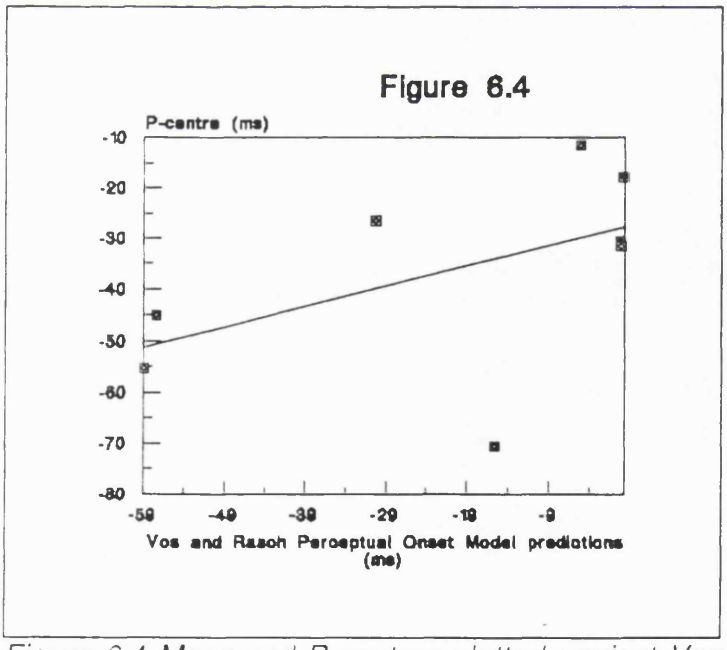


Figure 6.4 Measured P-centres plotted against Vos and Rasch perceptual onset model predictions

The regression line has the equation:

P-centre = - 27.8 + 0.393 Vos and Rasch prediction

The Vos and Rasch values are not significant predictors of the experimental P-centre values ($t_{1,6}$ = -1.43, p > 0.05).

The Vos and Rasch model of perceptual onset thus provides a poor predictor of the P-centres of these speech stimuli. The direction of the relationship, as shown in Figure 6.4 is reasonable; the earlier perceived onsets corresponding to earlier P-centres; but the relationship is not significant.

One explanation for the poor fit could be that the perceived onsets of these signals not co-occurring with the P-centres. Alternatively, the Vos and Rasch model may not be able to make predictions about speech data which contains fricative noise bursts such "t". Indeed, the model could simply be wrong. Whatever the reason, a significant amount of observed variance is not accounted for; it can be concluded that the Vos and Rasch model of Perceived Onset cannot account for P-centres in different speakers.

6.8.2 Centre of Gravity Model

Table 6.7 below shows the syllabic centre of gravity (Howell 1988 a&b) model predictions for the P-centre locations for the "one"s and "two"s stimuli.

Howell Model (ms)	"One"s	"Two"s
speaker SKS	-199.8	-182.3
speaker SHS	-130.9	-132.7
speaker LCE	-137.5	-150.1
speaker PH	-224.6	-118.3

Table 6.7 Syllabic centre of gravity (Howell 1988 a,b) model predictions for P-centre locations (ms)

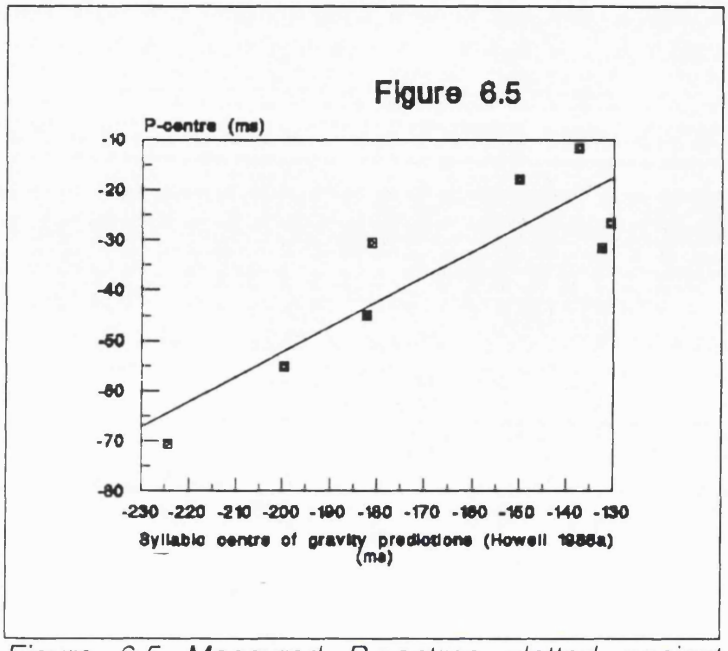


Figure 6.5 Measured P-centres plotted against Howell's syllabic centre of gravity model predictions

Figure 6.5 shows the calculated P-centres plotted against the syllabic centre of gravity predictions. The relationship looks roughly linear with later syllabic centre of gravity measures corresponding to later P-centres.

The observed P-centres were regressed against the syllabic centre of gravity predictions using linear regression with the equation fitted:

$$y = c + \alpha x$$

where y=P-centre and x=syllabic centre of gravity prediction

This plot has the regression equation:

P-centre = 35.1 + 0.446(syllabic centre of gravity)

The syllabic centre of gravity values are significant predictors of the experimentally measures P-centre values ($t_{1,6} = 4.23$, $p < 0.05$).

The syllabic centre of gravity - a representation of the intensity distribution over the entire signal - is a significant predictor of the observed variance in P-

centres. This model accounts for 73.0% of the observed variance. This shows that intensity changes are important in P-centre location.

6.8.3 Marcus's Model

Table 6.3 below shows the Marcus model P-centre predictions for the P-centre locations of the "one"s and "two"s stimuli.

Marcus model (ms)	"One"s	"Two"s
speaker SKS	-179.5	-140.6
speaker SHS	-146.7	-108.6
speaker LCE	-90.6	-121.8
speaker PH	-167.4	-142.8

Table 6.3 of Marcus model predictions of P-centre location (ms)

Figure 6.6 shows the calculated P-centres plotted against the Marcus model predictions. The relationship is roughly linear, with later Marcus model predicted P-centres corresponding to later calculated P-centres.

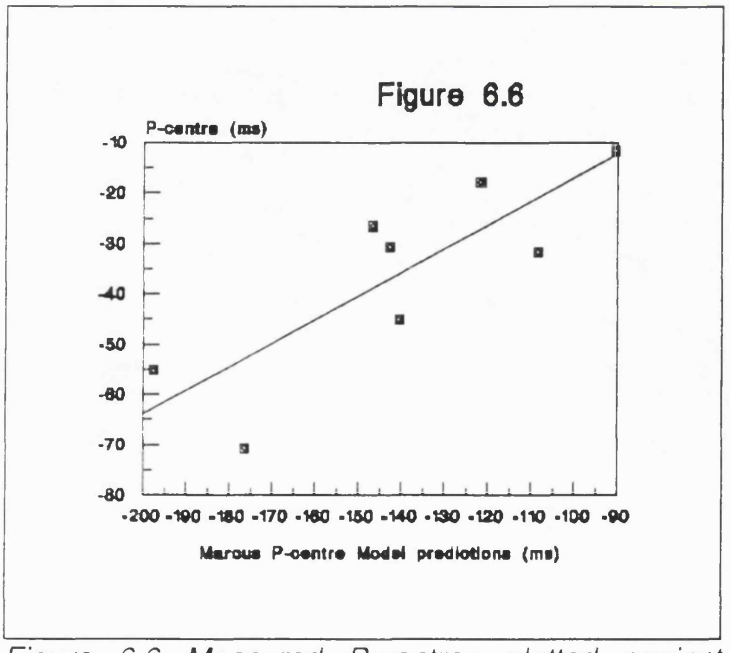


Figure 6.6 Measured P-centres plotted against Marcus's P-centre model predictions

The observed P-centres were regressed against the Marcus model predictions in a linear regression fitting the equation:

$$y = c + \alpha x$$

where y=P-centre and x=Marcus model prediction

This linear regression had the equation:

P-centre = 39.0 + 0.543 Marcus model prediction

The Marcus model values are significant predictors of the experimentally measured P-centres ($t_{1,6} = 3.98$, $p < 0.05$).

The Marcus model of P-centre location accounts for a significant amount of the variation in P-centres across speakers (67.9%). The relationship is in the same direction as that of Howell's model with P-centre location; for both models later calculated P-centres related to later predicted P-centres.

6.9 Implications for models of P-centre location

No single model emerges as the best predictor of P-centre location. Both the syllabic centre of gravity model and the phonetic model of P-centre location predict the experimental findings well. The Vos and Rasch model of perceptual onset does not account for any of the observed variation. In this section reasons why this was found will be discussed.

The Vos and Rasch model was not developed with speech sounds at all. They did carry out a dynamic rhythm setting task to measure differences between signals, and their experimental manipulation (rise time) is the same as that which Howell found altered the P-centres of speech and nonspeech sounds.

This model might be insufficient to account for the data because it is specific to non speech sounds. Alternatively, the use of a threshold relative to peak intensity might not be the best definition of onset characteristics. Certainly the model is sensitive to noise; the definition of the perceptual onset as the signal passing a threshold relative to the maximum intensity, with no explicit smoothing, leads to this sensitivity.

The Marcus and Howell models make very similar predictions about P-centre location. The slopes of the relationships are very similar (+0.543 and +0.496 respectively). Both of the slopes are different from 1.0, the slope that would be predicted by a direct 1:1 mapping between predicted and measured value. The intercepts are different, but that is because the P-centres, both predicted and measured, are relative values rather than absolute.

Conceptually, the Howell and Marcus model have some large differences, and some similarities. These will be considered attribute by attribute.

6.9.1 Global vs. local definition of P-centre

Both models relate P-centre location to attributes of the whole signal, rather than to single events within the signal. That is, they are both *global* models of P-centres, as opposed to *local* models.

6.9.2 Duration of the signal

Both models consider that entire duration of a signal contributes to the P-centre. However, Marcus model splits the duration into two portions, and weights the onset portion more heavily. The Howell model accords all the duration equal weighting.

6.9.3 Intensity changes

Marcus's model does not explicitly take into account the intensity time profile of a signal. Indeed, in his experiments, he found evidence which led him conclude that intensity changes have no effect on P-centre location (see Chapter 1). Howell's model relates the P-centre exclusively to the intensity variation of a signal. This variation is represented by a syllabic centre, which shifts with the energy distribution. However, Marcus's model does utilize information about intensity changes within a signal. This is via his definition of vowel onset as the peak increment in mid band spectral energy.

6.9.4 Spectral contents

Neither Howell's nor Marcus's model explicitly considers spectral attributes of a signal. Marcus's model does use spectral information as a model parameter, again in his definition of vowel onset. The mid band spectral energy lies between 500-1500Hz.

6.9.5 Specific to speech?

An attribute of a signal at a different level of analysis is whether it is a speech signal. Both Howell and Marcus state that their model's can be applied to speech and non speech, and are thus not speech specific.

Thus there are clear conceptual similarities between the two models (both are global, neither is specific to speech). There are large differences in the attributes stressed in each model. Marcus's model is principally concerned with the durations of different phonetic sections of a syllable, Howell's with the energy profile of a signal. Are they so different that they are orthogonal to one another in terms of the variance that they account for?

The two sets of predictions were correlated together using Pearson's *r*. The correlation of the two sets of predictions was significant ($r = 0.804$, $p < 0.02$, df

= 6). Correlation of the two sets of predictions would be expected to some degree, since they both predict P-centres significantly. This correlation is higher than would be expected if the two sets were to some degree orthogonal (the linear regressions indicated that the two models were not totally orthogonal). Taken with the similarities outlined above, it seems that, at least for these stimuli, the two models make very similar predictions.

The Howell and Marcus models will be considered again with different stimuli in the next experiments, to test whether they still make similar predictions. There is the possibility that a third factor, not explicitly considered in either model, covaries with each, and could be the best predictor of P-centre location.

6.10 Experiment 3b - Further speaker differences

Of the four speakers investigated in the previous experiment, three had participated in experiment two, producing rhythmic speech. In the previous experiment the P-centres of their utterances were compared to these produced intervals, to establish that differences in P-centre explained the timing variations.

There were four other speakers in experiment two; to extend the information on P-centre differences across speakers, and to relate differences in produced timing to P-centres. P-centres were calculated for examples of "ones" and "twos" for each of these four speakers.

6.11 Method II

The procedure described above in Method I resulted in P-centre values for all eight stimuli. P-centres were also calculated for a further eight stimuli - the "ones" and "twos" from SM DG WC and SR gathered in Experiment 2b, that is

all the other speech items not used in method I. Oscillograms of these speech items are shown in Figure 6.7. A reduced version of a dynamic rhythm setting task was used for these speech items, in order to gather further data.

6.11.1 Design II

The speech pairs of "one" and "two"s from SM DG WC and SR were set against each other and against the common reference sound in four dynamic rhythm setting tasks - a separate set of experimental blocks for each pair of stimuli. In each set, there were thus 3 stimuli and 9 experimental blocks of trials.

The subjects performed five trials in each block, seven blocks in each rhythm setting experiment (two blocks having been carried out in experiment 2), for four speakers' speech. The reference sound - reference sound settings were always identical and were therefore only carried out in two of the dynamic rhythm setting tasks to reduce the duration of the experiment. For each speaker' speech there were 9 sets of trials. All of the (one - two) and (two - one) pairings had been set in a dynamic rhythm setting task in the perception part of experiment 2; these were not therefore repeated; the data gathered in experiment 2 was used again. This was possible because the same three subjects who performed experiment 2 were used in this experiment, and the data gathering was contemporaneous.

The A - A interval was 1250ms, and the initial A - B interval was randomized.

The data from these dynamic rhythm setting tasks was treated separately from that data collected in Method I, due to the differences in the way that the P-centres of the stimuli were determined.

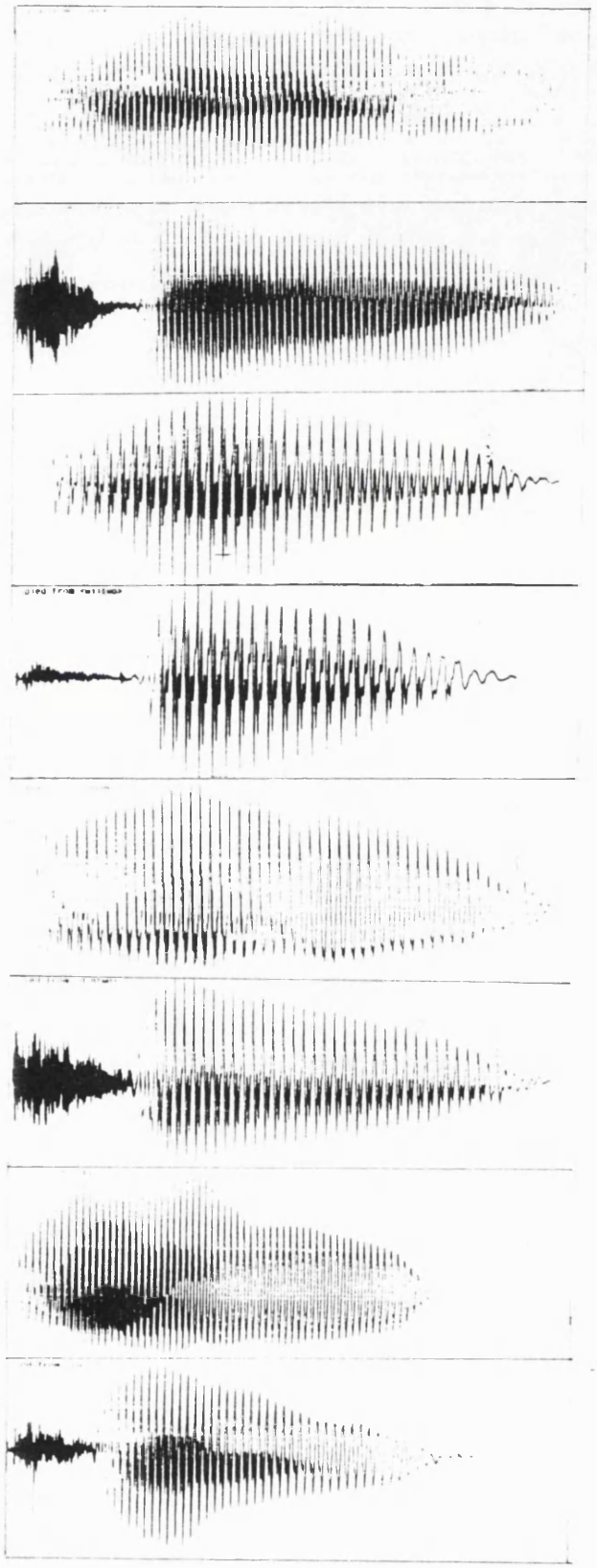


Figure 6.7 Oscillograms of examples of "ones" and "twos" from speakers SR, WC, DG and SM.

6.11.2 Subjects

Three subjects completed these experiments. Three of these were from the four subjects who participated in Method I.

6.11.3 Analysis

As described in the Design II section, the P-centres for these four speakers were not based upon the 5X5 matrix of trials as was done in the previous experiment. Instead reduced P-centre setting experiments were used. Each speaker was treated separately; their "one" "two" and the reference sound was used in rhythm setting task - a 3X3 matrix of stimuli. The intervals set in this experiment, plus the (one - two) and (two - one) interval from experiment 2, were used to calculate P-centres for each of the speech samples. The P-centres for each speaker have been determined only with respect to speech from that speaker and the reference sound; this reduces the degrees of freedom of the P-centre values.

6.12 Results II

The tables below show the mean final intervals for each of the rhythm setting tasks for each of the speakers. The shaded cells are the conditions where the same sound is set against itself (and the mean interval should thus be equal to [A - A]*0.5, that is 625ms).

In each of the four Tables 6.9, 6.10, 6.11, 6.12, the stimuli combinations where the same sound is set against itself (the shaded cells) show final intervals that are similar to 625ms. This indicates that the subject set intervals of near physical isochrony when setting identical sounds to a rhythm. Elsewhere in the

tables large deviations from isochrony can be seen, which implies that the subjects did not set these stimuli to physical isochronous rhythms in order to acheive perceptual isochrony. This indicates that the stimuli vary in their P-centres.

	"one"	"two"	ref
"one"	624.94 (9.48)	611.76 (11.00)	674.24 (25.27)
"two"	635.83 (17.96)	627.50 (11.50)	676.89 (28.30)
ref	571.28 (18.08)	579.82 (27.14)	624.22 (9.72)

Table 6.9 of mean final intervals in rhythm setting task - speech from SR

	"one"	"two"	ref
"one"	620.06 (10.47)	614.59 (13.13)	680.73 (16.99)
"two"	633.00 (15.66)	616.88 (10.61)	673.33 (33.25)
ref	570.82 (23.35)	581.89 (26.73)	

Table 6.10 of mean final interval settings in rhythm task - speaker WC

	"one"	"two"	ref
"one"	628.00 (13.27)	593.67 (17.74)	632.41 (19.21)
"two"	657.56 (19.50)	626.78 (14.90)	645.62 17.42
ref	614.67 (23.79)	597.94 (20.31)	623.78 (10.69)

Table 6.11 of mean final interval settings in rhythm task - speaker SM

	"one"	"two"	ref
"one"	619.71 (9.71)	612.29 (8.24)	659.29 (22.25)
"two"	639.06 (12.91)	624.4 (6.57)	655.24 (18.78)
ref	589.94 (16.37)	583.11 (19.86)	

Table 6.12 of mean final interval settings in rhythm setting task - speaker DG

The mean intervals set for each speaker were used in separate P-centre algorithm calculations; the P-centres were then adjusted such that the reference sound P-centre is equal to zero so that the P-centres can be compared across speakers, and with the P-centres from the previous experiment. Table 6.13 below shows these P-centres:

	P-centres (ms)	
	"one"	"two"
WC	-49.00	-52.00
DG	-31.67	-40.33
SM	-3.33	-29.67
SR	-43.00	-47.00

Table 6.13 P-centres for "ones" and "twos" from four different speakers - each speakers P-centres calculated independently, reference sound = zero.

The results show the basis for the differing one-two intervals across speakers in Experiment 2. For each speaker the one and two has a different P centre. There are large differences across speakers for the P centres of the single words one and two. These differences can be seen on Figure 6.8, which shows the different P-centres for each subject.

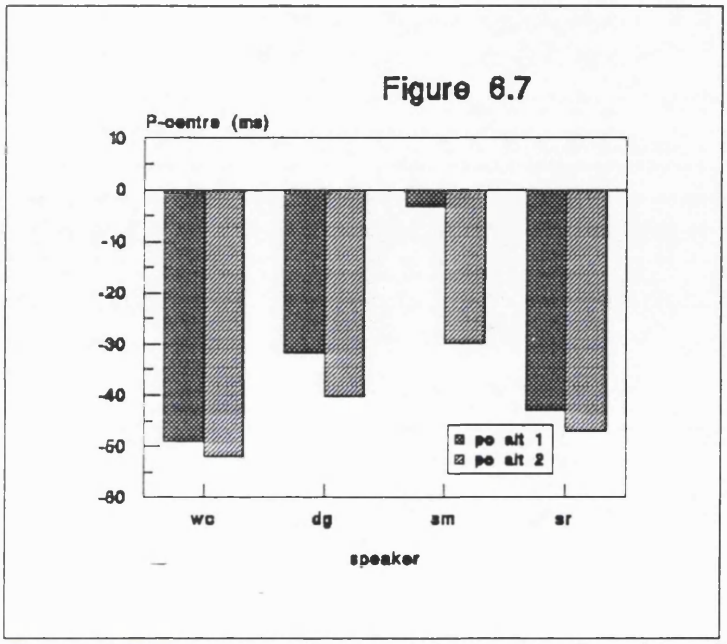


Figure 6.7 P-centres of speech items "one" and "two" from different speakers

6.13 Relating P-centres to production timing

These calculated P-centres were related back to the (one - two) intervals the four speakers produced in Experiment 2.

Speakers WC DG and SR produced larger (two - one) intervals than (one - two) intervals. This suggests that for these speakers the P-centres of the "ones" were early relative to the P-centres of the "twos"; to achieve perceptual isochrony the onset of the "twos" needed to be advanced nearer to the offset of the "one".

The P-centres of these three speakers confirm this hypothesis; Table 6.11 shows that for speakers WC SR and DG the P-centres of the "ones" are nearer to the onsets (closer to zero) than those of the "twos", regardless of the total size of each.

Speaker SM produced much larger (two - one) intervals than (one - two) intervals. This implies that the P-centre of the "two" is much later than that of the "one". The P-centres of the "one" and "two" from SM bear out this pattern (-3.33 and -29.67 respectively).

Therefore the P-centres from these four speakers, like the three described earlier, map well onto the timings they produce in perceptually even speech sequences.

6.14 Summary

In this chapter an experiment was described in which the P-centre for "ones" and "twos" for four different speakers were determined (by comparing all the "ones" and all the "twos" separately). These P-centres were congruent with the produced timings from three of the speakers in regular sequences. Implementations of the Vos and Rasch, Howell and Marcus models of perceptual occurrence were described, and used to make predictions which the measured P-centres were tested against. The Vos and Rasch model did not significantly predict the variance in P-centres; both the Marcus model and the Howell model predicted a significant amount of the variance, to a similar degree. The slopes of the two relationships was similar. Reasons for two such different models both predicting the P-centres were discussed. P-centre for four more speakers "ones" and "twos" were determined using rhythm setting tasks for each speakers' speech individually. The production timings for these speakers were compared to these P-centres, and again there was good mapping between produced timing and the P-centres of the utterances. This is evidence to support the suggestion at the end of Experiment 2 that the differences in production timing of perceptually regular sequences between speakers be based upon P-centre differences between the speakers.

Chapter Seven
Experiment four:
Does Infinite peak clipping alter the P-centre of speech signals?

Abstract
This experiment addressed Tuller and Fowler's 1981 finding that infinitely peak clipping speech items (that is, rendering invariant the acoustic profiles of the speech sounds) does not affect the P-centres of the stimuli. The aim was to determine whether infinite peak clipping does affect P-centres if care is taken over the stimuli and the experimental paradigm. Infinite peak clipping was performed digitally, care was taken over the reproduction of the signal at the subject's ear, and a dynamic rhythm setting task was used. The infinite peak clipping altered the P-centres of the speech items, in each case moving the P-centres nearer towards the onset of the sound. The implementations of the P-centre models described in Experiment three were used to generate predictions about the P-centres of the stimuli; these were compared to the experimental results. Contrary to expectations, the acoustic model (Howell 1988 a&b) did not account fro the data, while the phonetic model of Marcus (1981) made more accurate predictions. Reasons for this result are discussed.

7.1 Introduction

The results of Experiments Two and Three show that different speakers produce different patterns of intervals between speech signals when speaking rhythmically, and that these differences are due to the different P-centres of their produced speech. These P-centre differences are predicted by two current models of P-centre location (Howell 1988a, Marcus 1981). A tentative conclusion is that this correspondence between two very different models is that the P-centres are varying with some attributes of the intensity/time profiles of the signals, for instance the onset characteristics.

Previous work, as mentioned in the introduction, has provided empirical evidence concerning the alteration of intensity variation of speech and nonspeech sounds over time and the effect on P-centre location. Such evidence has been used to discriminate between the amplitude / time, and direct perception approaches. There are several amplitude / time models in the literature, which account for perceptual rhythm judgements in terms of the amplitude / time characteristics of a signal. These are described in more detail in Chapters One and Two.

In developing a simple acoustic model Howell (1984, 1988a) varied the P-centre of signals by manipulating the amplitude / time profile; he found that altering the rise-time of speech signals shifted the P-centre location. He proposed a simple amplitude envelope model of P-centre location which accounts for the energy characteristics at onset and offset, and also duration effects.

Vos and Rasch (1981) quantified the energy distribution at the onset of a signal using a simple threshold model, and related this to the perceptual onset of a musical tone (analogous to the P-centre for signals with a rapid rise time). The slower the rise time of a signal, the later the intensity profile crosses the threshold, and so the later the perceptual onset occurs.

Gordon (1987) modelled the perceptual attack time of musical notes (which he explicitly compared to P-centres in speech). His model also incorporated onset amplitude characteristics, but instead of a simple threshold, he used the rise-time or attack rate of the signal. This was based on experimental results, which showed that for signals with a rapid rise-time, the perceived attack time was influenced by amplitude characteristics; when the rise-time was slower, the perceived attack time was influenced by spectral cues.

The models of Vos and Rasch, Gordon and Howell were all based on experimental data from explicit manipulation of the amplitude / time distribution at the onset, although the models vary in whether they conceive of P-centres as global or local phenomena.

Other experiments have been conducted, which aim to support different models of P-centre location. Tuller and Fowler (1981) provided evidence that the amplitude envelope of a speech sound did not affect its P-centre. They used a technique called 'infinite peak clipping', which removes amplitude/time variations from a signal. To infinitely peak clip a signal means altering it so that all positive sample values become equal to a maximum, all negative values equal to a minimum, and all zero values remain the same. This results in a square waveform, with a rectangular amplitude envelope. All the frequency information is retained, and the speech is harsh and noisy, but still intelligible.

They found that infinite peak clipping did not affect listeners' judgments of the 'naturalness' of perceptually isochronous sequences. They hypothesised therefore that the P-centres of the sounds had not been affected by the distortion of the amplitude envelope of the speech sounds.

Tuller and Fowler used however an imprecise measure of perceived timing (Howell 1988a, Pompino-Marschall 1991). Subjects had to choose which of one naturally timed and one altered sequence sounded more 'natural'. Subjects chose the sequence which was naturally timed, whether or not the speech sounds were infinitely peak clipped. There were, however, only two different timing conditions.

7.2 Tuller and Fowler's experiment

Fowler and her colleagues (Fowler 1979, Tuller and Fowler 1980) consider that the regularity in timed speech activity occurs in production with respect to activity in certain muscle groups. Thus, perceptual adjustments would not be expected to align with respect to any acoustic referent. This perspective predicts no simple relationship between the articulation of a sound, and the acoustic structure of the produced sound. Tuller and Fowler (1981) infinitely peak clipped speech sounds to rule out amplitude as a determinant of P-centre location (see Method for details of infinite peak clipping). Energy distribution could be dismissed as a determinant of P-centre location, if gross distortion, of this parameter caused no shift in P-centre location. More specifically, energy time variation could be ruled out of removing all the variation had no effect. They found that infinitely peak clipping speech sounds did not shift their P-centres. Listeners rated naturally timed sequences as more natural, rather than physically isochronous sequences, whether or not the speech was infinitely peak clipped.

In Fowler, Whalen and Cooper (1988) oscillograms of one pair of syllables Tuller and Fowler (1981) employed are presented (shown in figure 7.1). These stimuli are not infinitely peak-clipped. Instead, some specified portions seem to have been scaled up to use the full range of their D-to-A converter. This process would introduce high frequency noise by raising the noise floor, which may have distorted the spectral qualities of the signal. There are also differences in rise time at syllable onset. At the end of the syllables, the non-clipped sounds are falling in amplitude, while the clipped ones rise. There are durational differences between the peakclipped and non-peak clipped syllables. This could have affected the P-centres of the signals.

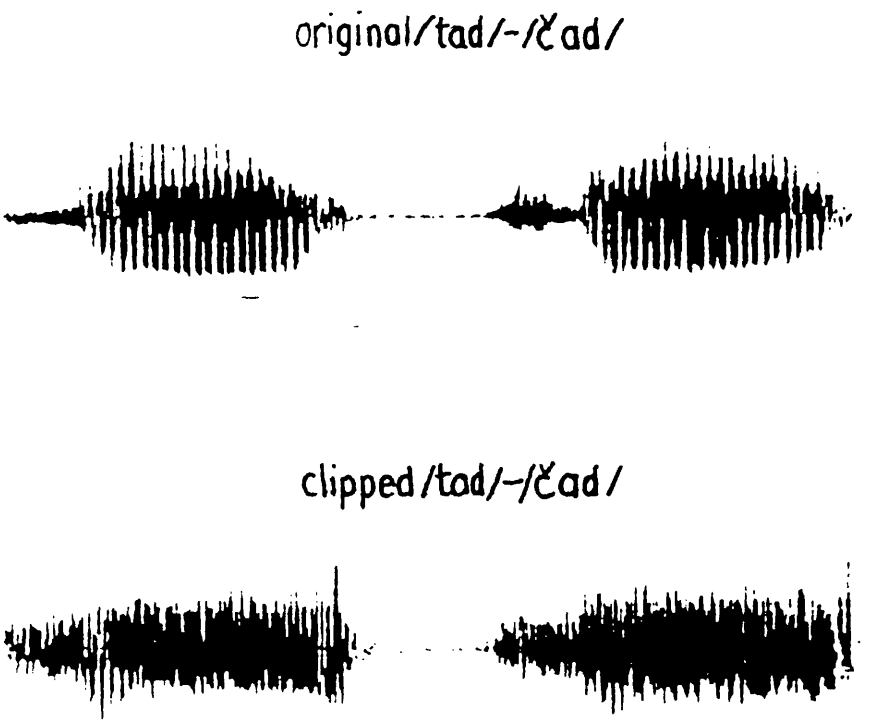


Figure 7.1 Oscillograms of a pair of normal and 'infinitely peak clipped' stimuli used by Tuller and Fowler 1981, presented by Fowler et al 1988 (reproduced with permission).

This experiment has been interpreted as showing no shift in P-centre (Tuller and Fowler 1981, Cooper et al. 1986). Tuller and Fowler did not, however, test for such a shift very rigorously. 'Infinitely peak clipped' and not peak clipped sequences were presented to subjects; the intervals between the syllables were either acoustically equal (physical isochrony) or corresponded with those which sounded evenly timed when spoken naturally (perceptual isochrony). On a forced choice task to choose which sequences of syllables sounded more natural, subjects chose the sounds separated by the naturally produced interval. This was not a satisfactory number of levels of the independent variable (physical intervals). What would the results of been if they had used several different interstimuli intervals that varied regularly between the two extremes? To determine the P-centres of the stimuli and thus test the experimental hypothesis sufficiently, the subjects would have adjusted infinitely peak clipped and natural syllable pairs in a rhythm setting task designed to locate the P-centres.

Referring back to the infinite peak clipping experiment of Tuller and Fowler (1981), Fowler et al (1988) acknowledge that the signals do not display the rectangular amplitude envelope which truly infinitely peak clipped stimuli would show; they attribute the lack this to the deemphasizing filters at output on the Haskins PCM system. Tuller (personal communication) has conceded that the use of the more precise rhythm setting method might well result in a shift in P-centre caused by infinite peak clipping becoming apparent.

There is thus an apparent contradiction between the infinite peak clipping experiment and the ramping experiments mentioned earlier. Tuller and Fowler claimed that infinite peak clipping did not alter the P-centre location, whilst the results of the ramping experiments suggest that this manipulation of the amplitude/time distribution would affect P-centre location. Specifically:

i) Vos and Rasch's and Gordon's models would predict that infinitely peak clipping would shift the P-centre towards the onset of the signal, by decreasing the rise-time of the stimuli.

ii) Howell's Centre of gravity account would predict that the P-centre would shift along with the syllabic centre of gravity - the point in time at which half of the energy has been integrated. The predicted P-centres would thus be different for the infinitely peak clipped stimuli, due to the large amplitude / time change caused by this manipulation. The direction of P-centre shift depends on the original stimuli.

This experiment will investigate whether infinite peak clipping a signal alters its P-centre location when using a more sensitive experimental paradigm than that of Tuller and Fowler (1981), and correctly infinitely peak clipped speech. In addition, whether any shift would be predicted by any other model of P-centre location.

7.3 Method

7.3.1 Stimuli

Four speech samples ("la", "ra", "wa", "ya") were spoken by an adult male British-English speaker. They were recorded on DAT, and then digitized at 20kHz. Figure 7.2 shows the oscillograms the stimuli.

To avoid the problems inherent in using direct amplification of the waveform to acheive peak clipping software was written to infinitely peak clip the speech tokens (SEE APPENDIX). In the program, each positive sample was increased to a maximum, and each negative sample value amplified to a minimum.

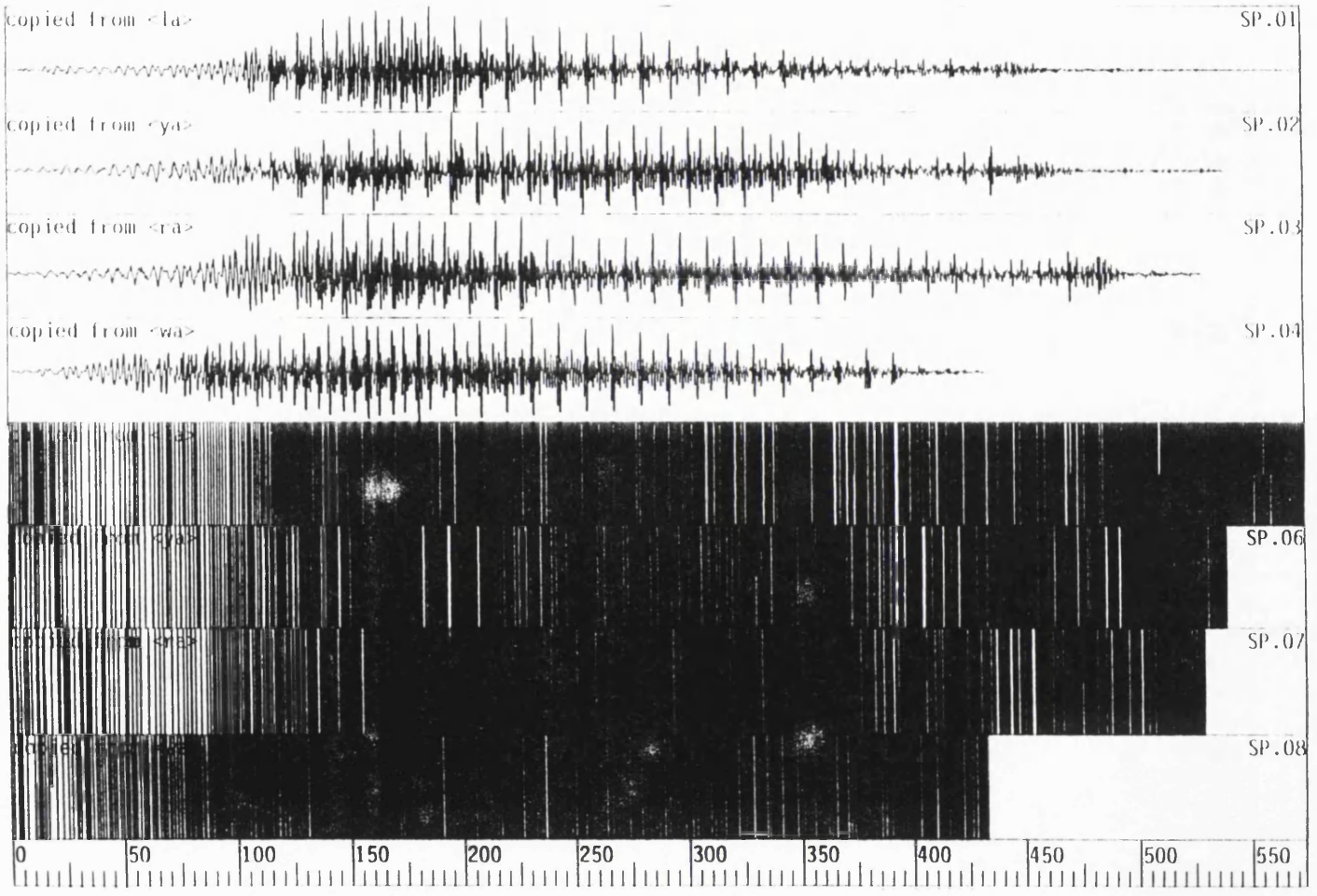


Figure 7.2 Oscillograms of normal speech stimuli "la, ya, ra, wa" and their infinitely peak clipped equivalents.

The original speech items are referred to as speech stimuli, and the infinitely peak clipped speech as $speech_{ipc}$ stimuli.

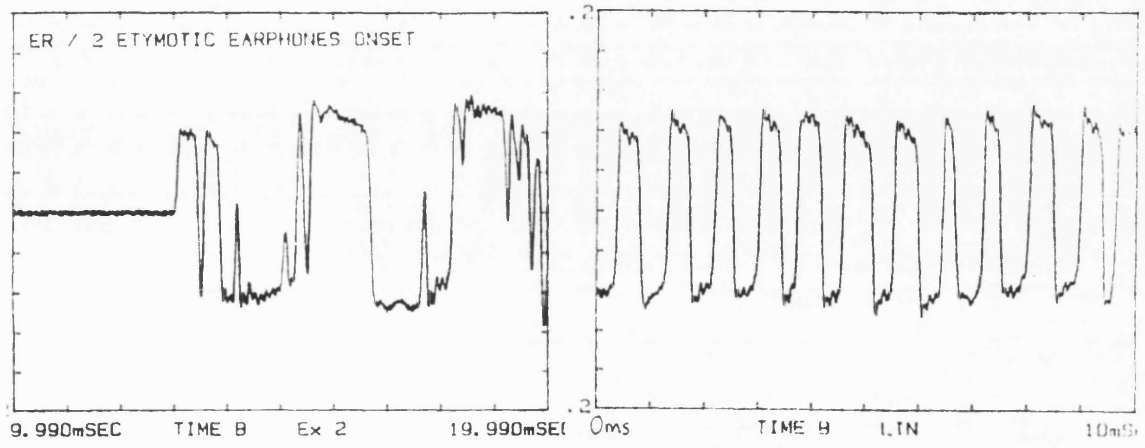
7.3.2 Apparatus

The stimuli were played to the subjects using the apparatus described in Chapter 4. The sound was delivered to the subject via an ER-2 insert earphone, designed for a flat frequency response at the ear. It was calibrated for phase and amplitude response and showed no distortion. The construction of the stimuli is unimportant if the equipment used in the experiment did not reproduce the signal correctly at the ear, since amplitude and phase distortion of the signal by the equipment would radically distort the stimuli. To ensure therefore that the square waveform and rectangular amplitude envelope which had been constructed were reproduced at the subjects' ear, all the playback equipment used was calibrated for phase and amplitude distortion at the frequencies needed. The calibration was performed using a 0.6cc coupler, and an ONO SOKKI real time Fourier spectral analyzer. The amplifier (a QUAD 520 Power amplifier) and filter (Barr and Stroud variable FF3) used were checked in this way, and only equipment which did not distort the signal was used.

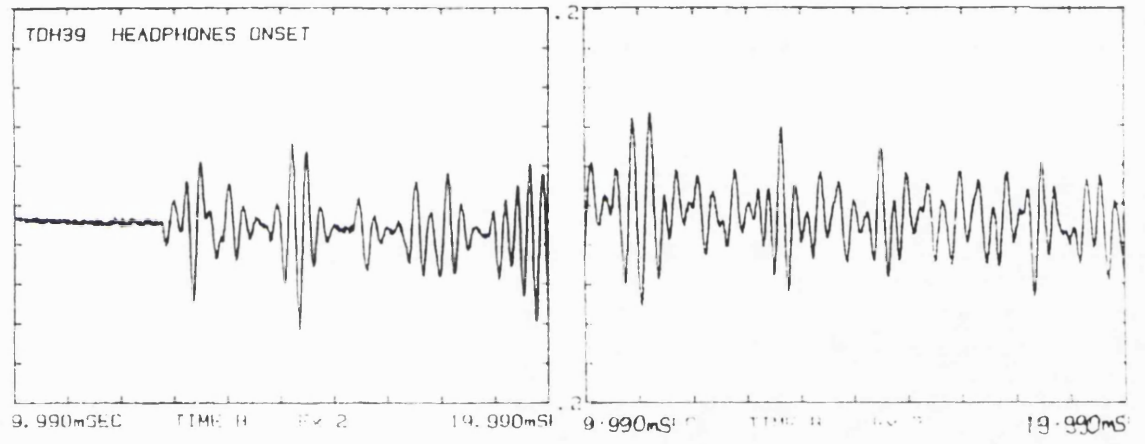
The square waveform put into the system was therefore not altered. Figure 7.3.1 shows the amplitude/time plot of the output of the system playing out some infinitely peak clipped speech; an example of the onset of a peak-clipped signal, and a middle portion of the same signal are shown.

To illustrate the effect of a non-calibrated headphone, Figure 7.3.2 shows the onset and a middle portion of the same signal played through the same system, but with a standard TDH-39 headphone instead of the ER-2 insert earphone. The distortion leads to a signal with a ramped onset, a non-rectangular amplitude envelope, and a non-square waveform.

7.3.1



7.3.2



Figures 7.3.1 & 7.3.2: Amplitude/time plots of the same squarewave signal played through an ER-2 insert earphone (7.3.1), and a THD-39 headphone (7.3.2)

7.3.3 Design and Procedure

P-centres were determined using the dynamic rhythm setting task described in Chapter 4. The aim of the experiment was to test whether there is a difference in the P-centre of a speech token (eg. "ra") and its infinitely peak clipped version (eg. "ra_{ipc}"). A preliminary experiment was run, in which a speech token was set directly against its infinitely peak clipped equivalent ($speech_{ipc}$) in a full dynamic rhythm setting P-centre determining task. If there was no difference

in the P-centres of the two, then perceptual isochrony would have been physical isochrony (physical isochrony meaning physical onset to onset isochrony). Any deviation from physical isochrony would indicate a shift in P-centre due to the manipulation.

This study indicated that the direct comparison of the speech against the $speech_{ipc}$ was too difficult for the subjects. Making a rhythm setting judgement involves streaming together the component signals (Seton 1989), that is, to make a setting, the subjects need to hear the sequence of sounds as a coherent stream of acoustic information (Bregman and Campbell 1971). Subjective reports indicated that it was hard for the subjects to hear the speech and the distorted speech as part of the same perceptual sequence, and this led to problems in determining whether the sequences was isochronous or not. This was reflected in the experimental results. Instead of a full P-centre determining task, therefore, each speech token used in the experiment was set to a rhythm against the reference noise (50ms noise). Any difference in interval set between a speech token and the reference noise, and the equivalent $speech_{ipc}$ and the reference noise, would indicate a shift in P-centre location caused by the infinite peak clipping. Over the course of the whole experiment, each token was set as both A and B in the rhythm setting task (as described in Chapter 4). As there were nine tokens (four speech tokens, four $speech_{ipc}$ tokens, and one reference noise) there were therefore eighteen combinations. Each combination was presented to a subject four times. The order of trials was randomized.

7.3.4 Subjects

Four subjects (one female, three males) took part in the experiment. All had taken part in rhythm setting experiments before and were well practised. All were naive to the aims of the experiment.

7.4 Results

Table 7.1 below shows the mean intervals set for all the stimulus pairings. It can be seen that there are differences between the intervals set for normal speech, and the peak clipped equivalent speech. The intervals set for the peak clipped speech are smaller, indicating that the P-centres of the peak clipped stimuli are nearer to the onsets of the signals.

Stimuli Pairings	
"la" - ref	ref - "la"
702.25 (22.94)	571.67 (16.84)
"ya" - ref	ref - "ya"
681.30 (36.40)	578.55 (21.56)
"ra" - ref	ref - "ra"
713.33 (25.02)	555.42 (27.82)
"wa" - ref	ref - "wa"
654.50 (30.84)	588.08 (13.81)
"la"$_{ipc}$ - ref	ref - "la"$_{ipc}$
680.42 (28.02)	587.67 (27.41)
"ya"$_{ipc}$ - ref	ref - "ya"$_{ipc}$
637.42 (15.63)	599.58 (19.95)
"ra"$_{ipc}$ - ref	ref - "ra"$_{ipc}$
662.80 (35.40)	587.90 (39.30)
"wa"$_{ipc}$ - ref	ref - "wa"$_{ipc}$
630.50 (15.51)	611.67 (12.06)

Table 7.1 means and SD's of final intervals set for each stimulus pairing

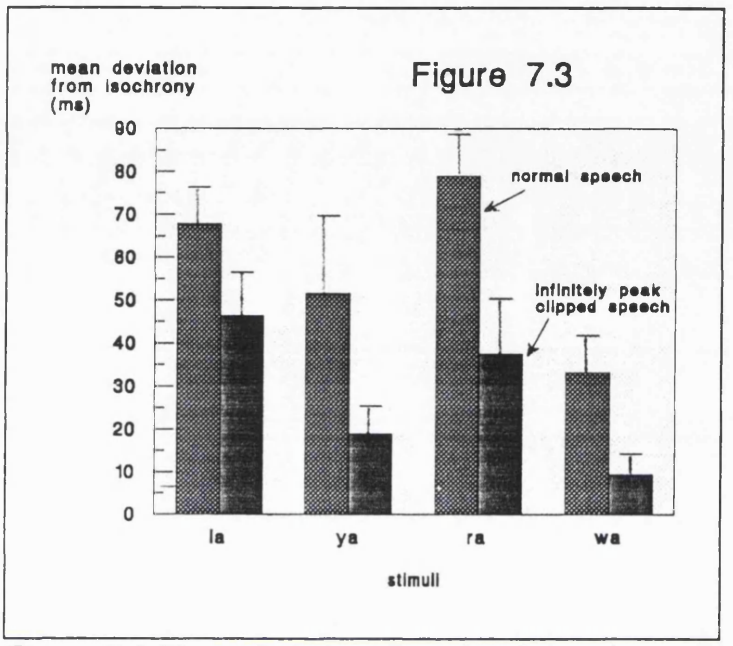


Figure 7.4 Mean deviations from isochrony for each speech and speech$_{ipc}$ stimulus

From Table 7.1, the settings made with the infinitely peak clipped stimuli look to vary less from physical isochrony than those set with the normal speech. This implies that the P-centres of the infinitely peak clipped speech have shifted. The mean absolute deviations from isochrony are shown in Table 7.2. The standard deviations are large in some cases.

	mean deviations (ms)	
	speech	speech$_{ipc}$
"la"	67.79 (24.61)	46.37 (28.64)
"ya"	51.57 (29.98)	18.92 (18.74)
"ra"	78.98 (27.59)	37.46 (36.58)
"wa"	33.21 (23.67)	9.42 (14.17)

Table 7.2 absolute mean deviations from isochrony and SD for each stimulus (ms)

Figure 7.4 on the previous page shows these mean intervals graphically; for each speech item the normal and the peak clipped intervals can be compared. The differences between the mean intervals of speech and the $speech_{ipc}$ in Figure 7.4 are large, although they do vary across stimuli. In each case the $speech_{ipc}$ deviations from isochrony are *smaller* than the normal speech items.

The larger standard deviations, and the degree of variation are found for the infinitely peak clipped stimuli, and reflect the difficulties the subjects reported when setting these stimuli to a rhythm due to their (reported) unpleasant qualities. To test the significance of these deviations, the data was transformed using a square root function. This reduced to overall mean and standard deviation from 44.53ms (31.73) to 6.150ms (2.597).

The statistical significance of these results was analyzed by examining the mean absolute deviations from isochrony of the intervals set for each stimulus pairing. As described in previous chapters, this entails subtracting the [A - A]*0.5 value from each interval (that is interval - 625ms). This gives a measure of the amount of deviation from physical isochrony needed to set the stimuli to a perceptually even rhythm.
A multiple linear regression was performed fitting the equation:

$$y = c + \alpha(x_1) + \beta(x_2)$$

where y=deviation from isochrony, x1=speech item, x2=condition

This was to test whether the speech items, the condition (whether the speech was normal or infinitely peak clipped) or the subjects were significant predictors of the deviations from isochrony. The predictors were thus speech item ("la" "ya" "ra" or "wa"), condition (peak clipped or not) and subject.

The regression had the equation:

transformed absolute deviations from isochrony = 11.4 - 0.658(speech item) - 2.42(condition)

The predictor **speech item** was significant ($t_{1,187}$ = -4.65 $p < 0.05$).

The predictor **condition** was significant ($t_{1,187}$ = -7.61, $p < 0.05$).

The predictor **subject** was not significant ($p > 0.05$).

To test for subject differences, the predictor **subjects** was entered as a dummy variable (i.e.. variable $\gamma(x3)$ entered into the above model instead of variables x1 and x2). The subject 2 was a significant predictor of the intervals set; (due to her setting larger deviations than the other three subjects) ($t_{1,189}$ = 4.13, $p < 0.05$). The other subjects were not significant predictors ($p > 0.05$). The predictor subject 2 accounted for 8.3% of the observed variance, while the predictors speech item and condition accounted for 29.8%. Subject 2 may have differed due to her being an ensemble music performer, although she had been a subject previously and had not shown differences from other subjects.
Both the speech item and the condition were accounting for significant amounts of variance in the absolute deviations. This conforms the observations made about the mean intervals shown in Table 7.1; there are differences between the speech items and the infinite peak clipping alters the intervals set for all of the stimuli. Figure 7.4 shows the mean absolute deviations for all the stimuli. Again, as noted earlier, infinitely peak clipping the speech sounds reduces the absolute deviations from isochrony - that is shifts the P-centres of the speech sounds nearer to the onsets of the signals.

7.5 Conclusions

Tuller and Fowler (1981) results were not replicated. Infinite peak clipping a speech stimulus changes its P-centre location. Changes to the amplitude envelope of a speech signal do affect its P-centre.

The change in P-centre location caused by infinitely peak clipping the speech stimuli always results in a *smaller* mean deviation from isochrony than that

found for the normal speech; this indicates that P-centre for the infinitely peak clipped speech is nearer to the onset of the signal. The P-centre has shifted forwards in time. This result is predicted by models, which predict that the shorter the rise time, the earlier the P-centre occurs (Gordon 1987, Vos and Rasch 1981).

The results therefore show that the amplitude profile of a speech signal affects the P-centre of that signal. Which acoustic model best predicts this result? In the next section the two P-centre models and one perceptual onset model are tested as predictors of the measured P-centres (deviations from isochrony).

7.6 Implications for the Articulatory model of P-centre location

The P-centres of speech sounds are affected by the energy profile of the signal. Altering this parameter by infinite peak clipping affects the P-centres of the speech sounds. This finding is utterly contradictory to the articulatory account of P-centre location solely because this particular definition places P-centres within a class of speech specific, directly perceived phenomena. This manipulation, which preserves the 'underlying articulatory gestures' (that is frequency content) should not therefore shift the P-centres. This result means that an articulatory account of P-centre location must incorporate the theoretical concept that articulatory gestures are at least expressed in the acoustic waveform in some way, thus if the acoustic signal is altered, perceptual changes occur. A personal communication from Whalen (a coauthor of the 1988 paper) suggested that the infinite peak clipping manipulation of the speech sounds might cause phonetic changes which would alter the P-centres. This is incorrect on two counts. Infinite peak clipping does not alter the phonetic identity of speech sounds; the speech remains intelligible. Even if this manipulation did cause phonetic changes this would not *per se* lead to P-centre

shifts, as, Cooper, Whalen and Fowler (1986) demonstrated that phonetic identity and P-centre location did not covary.

7.7 Predictions of acoustic models

In Experiment Three, three models of P-centre location were tested against the P-centres for speech from different speakers. Two of the models (Howell and Marcus) predicted the variance well. These three models will be tested against the mean deviations from isochrony for the stimuli in this experiment, to examine which best fits the data, and distinguish between the accounts further. The implementations of the models are described in Experiment Three.

7.7.1 Vos and Rasch's model

Vos and Rasch's Perceived onset model would predict that, since the onset is perceived when the intensity passes a threshold 15dB below peak intensity level, the onsets of the stimuli would move forward in time when the speech stimuli were infinitely peak clipped. Indeed, since all the physical onsets would be identical after this manipulation, the perceived onsets of these stimuli would be the same. Table 7.3 below shows the Vos and Rasch model predictions for the stimuli.

	speech	$speech_{ipc}$
"la"	-96.8	0.00
"ya"	-67.0	0.00
"ra"	-57.0	0.00
"wa"	-41.2	0.00

Table 7.3 Vos and Rasch model predictions (ms)

The observed deviations are plotted against the Vos and Rasch model predictions in Figure 7.5. The deviations were regressed against these predictions fitting the equation

$$y = c + \alpha x$$

where y=deviation from isochrony and x=Vos and Rasch model prediction

The regression equation is
deviation from isochrony = 28.1 - 0.453(Vos and Rasch model prediction)

The Vos and Rasch model values were a significant predictor of the observed variation ($t_{1,6} = -2.70$, $p < 0.05$).

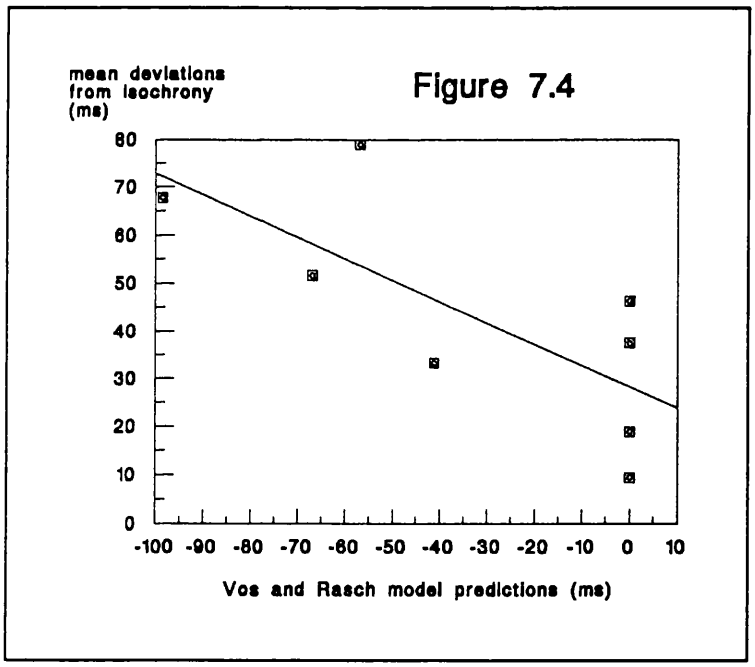


Figure 7.5 Mean deviations from isochrony plotted against Vos and Rasch perceptual onset predictions

The Vos and Rasch model therefore accounts for a significant proportion of the observed variance. The plot on Figure 7.5 however indicates that the model does not account for the speech$_{ipc}$ deviations - it predicts all the deviations

should be equal to zero for these stimuli. It provides a reasonable fit for the normal speech.

7.7.2 Howell's model

The centre of gravity model (Howell 1984, 1988a&b) would predict that the P-centre would vary with the syllabic centre of gravity. The P-centre is thus a result of the energy distribution over the whole signal. The syllabic centres for all the stimuli, peak clipped and non peakclipped, and are shown on Table 7.4 below:

	speech	$speech_{ipc}$
"la"	-196.2	-284.0
"ya"	-245.5	-268.3
"ra"	-209.4	-263.0
"wa"	-186.1	-215.4

Table 7.4 Howell model P-centre predictions (ms)

Figure 7.6 shows the observed deviations plotted against these predicted values.

The deviations were regressed against these predictions fitting the equation:

$$y = c + \alpha x$$

The regression line has the value:

deviation from isochrony = 79.2 + 0.155 syllabic centre of gravity

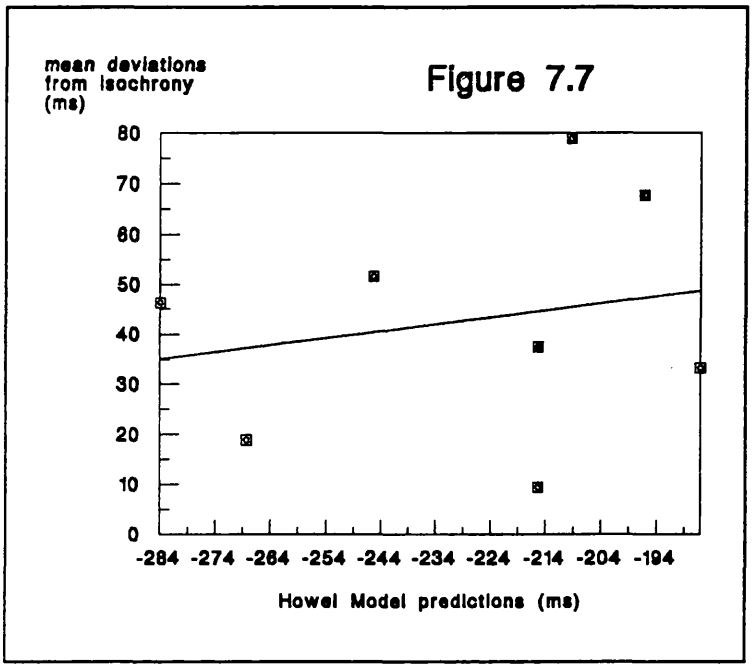


Figure 7.6 Mean deviations from isochrony plotted against Howell's syllabic centre of gravity model predictions

The syllabic centre of gravity values are not significant predictors of the experimentally observed variation ($p > 0.05$).

The syllabic centre of gravity model does not therefore account for a significant amount of the variance. This is because this model predicts a shift of the P-centre *away* from the onset when the speech is infinitely peak clipped (see Table 7.4 above). Instead, as was noted earlier, the P-centre of every speech item is shifted *towards* the onset of the signal by this manipulation.

7.7.3 Marcus's model

The implementation described in experiment 3 of Marcus's model was used to derive predictions of P-centres for the stimuli. These are shown in Table 7.5 below.

	speech	$speech_{ipc}$
"la"	-224.3	-201.2
"ya"	-212.7	-194.7
"ra"	-222.3	-132.3
"wa"	-175.3	-108.3

Table 7.5 Marcus Model P-centre predictions (ms)

The measured deviations from isochrony were plotted against these values and shown in Figure 7.7.

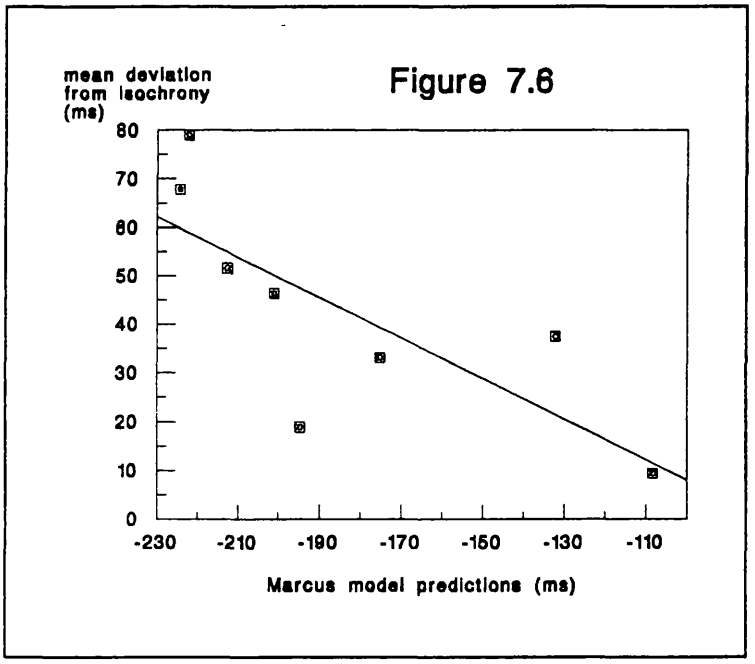


Figure 7.7 Mean deviations from isochrony plotted against Marcus model P-centre predictions

The deviations were regressed against these predictions fitting the equation:

$$y = c + \alpha x$$

where y=deviations from isochrony and x=Marcus model predictions

The regression line of this plot had the equation:

deviation from isochrony = - 33.7 - 0.417 Marcus model prediction

The Marcus model values are significant predictors of the experimentally observed variation ($t_{1,6} = -2.89$, $p < 0.05$).

Marcus's model therefore accounts for a significant amount of the variance observed in the deviations from isochrony (51.1%). The relationship is positive - the later Marcus's model predicts the P-centre, the larger the deviation from isochrony of the signal.

7.8 Discussion.

The Vos and Rasch model is a significant predictor; as discussed above however it does not account for the P-centres of the $speech_{ipc}$ stimuli. This is because the Vos and Rasch model is noise sensitive; since the initial amplitude value of the $speech_{ipc}$ stimuli is the same as the maximum amplitude the model predicts that the perceived onset is instantaneous. This is due to their quantification of onset characteristics is terms of a threshold of intensity, relative to peak intensity. This may not be the best representation of onset effects and does not deal well with stimuli such as the $speech_{ipc}$, which have 0ms rise time. This model does provide a reasonable fit for the normal speech stimuli, which indicates that for certain speech sounds it may be possible to model the P-centre with reference to the onset intensity characteristics.

The model (syllabic centre of gravity) which uses amplitude variation over the entire duration as the model parameter, and would therefore predict that infinite peak clipping affects the P-centres of speech does not account for a significant amount of the variance in deviations from isochrony.

For the Howell model the reasons for the lack of significance are clear. The model relates P-centres to the distribution of energy over the duration of the entire signal. An increase in energy at the onset shifts the P-centre towards the onset of the signal; an increase at the offset shifts the P-centre towards the offset of the signal. If the energy were increased over the whole signal by an equal amount, there would be *no change in the P-centre of the signal.*

Infinite peak clipping involves amplifying the energy of a signal to a maximum amplitude over the whole signal duration. It does not therefore affect the energy of a signal at the offset or onset only; nor does it increase the energy of the whole signal by a similar amount. Instead it causes a varying amount of energy change over the whole signal depending on the original energy variation over time. Thus if the onset of a signal has a very low amplitude, then this section will have its amplitude very much changed by infinite peak clipping. If the onset had an amplitude profile high in energy, then that section of the signal would not be much changed by infinite peak clipping. The point of this is that infinite peak clipping causes different amounts of change to the amplitude distribution of a signal according to what that distribution originally was. For the speech items used in this experiment the maximum change in amplitude occurred towards the offset of the signal. This can be seen in the comparison of the speech and $speech_{ipc}$ oscillograms in Figure 7.2. In each case the normal speech drops in amplitude soon after onset, and after infinite peak clipping the offsets are greatly increased in energy. This observation is quantified by the *syllabic centres of gravity of these stimuli.* The shift in syllabic centre of gravity indicates the region of the speech signal where there has been the most change in the amplitude profile caused by the manipulation. That is, the Howell model predictions indicate the alteration of the energy distribution caused by this manipulation.

The P-centres of the infinitely peak clipped stimuli are not therefore influenced by the increase in energy towards the offsets of the stimuli; instead the changes at the onset of the signals appear to influence the P-centre locations.

This factor - energy distribution at the onset - is precisely that which Marcus's model would state has no effect on P-centre location. His model relates P-centres to the durations of portions of the signal - the very parameters not affected by infinite peak clipping. Yet his model provides the best fit for the data, if not a perfect fit. This is because his model relies on the durations of different portions of the syllable - distinguished by vowel onset. His definition of vowel onset is the peak increment in mid band spectral energy (500-1500Hz). It is this variable that is affected by the infinite peak clipping. Infinite peak clipping shifts the 'vowel onset' forward and thus shifts the predicted P-centre towards the onset of the signal.

7.9 Summary

Measured by mean deviation from isochrony when set against the reference sound, there is a difference in the P-centres of the speech items "la" "ya" "ra" and "wa". If these speech items are infinitely peak clipped, the P-centres of all these items are altered. The P-centres shift towards the onsets of the speech sounds. Tuller and Fowler's (1981) finding was thus not replicated. A purely articulatory account of P-centre location is not sufficient to explain P-centre location; acoustic parameters such as the energy profile affect P-centre location. The shift in P-centre was significantly predicted by the Vos and Rasch perceptual onset model (although this model could not account for the $speech_{ipc}$ stimuli; Howell's syllabic centre of gravity model was not a significant predictor of the observed deviations. The P-centres were not apparently affected by the energy increase over the entire signal; the changes at the onset were the main determinant. The Marcus model of P-centre location did predict a significant

proportion of the observed variance; this was due to the alteration of 'vowel onset' as defined by this model by the infinite peak clipping.

Chapter Eight
Experiment five:
Amplitude envelope and P-centre location in perception and production of rhythmic speech

Abstract
The effect on P-centre location of increasing the amplitude of a syllable final "t" burst was examined, in both perception and production. This was to test the prediction of the syllabic centre of gravity model of P-centre location (Howell 1988 a,b) that alterations in any part of the amplitude envelope will affect the P-centre of the signal. Therefore the "t" burst increase at the offset of the signal should shift the P-centre towards the offset of the signal. The experimental results indicated that this manipulation does not affect the P-centre of a syllable either in production or perception.

8.1 Introduction

Experiments Two and Three showed that there are differences between speakers in the P-centres of their utterances. The phonetic model of Marcus, and Howell's syllabic centre of gravity model both provided good predictions of this data. In Experiment four, infinite peak clipping was found to alter the P-centres of speech sounds. Marcus's model was a reasonable fit for these results; Howell's model was not a good predictor of the observed variance.

There is thus a discrepancy; the energy distribution of a signal is affecting the P-centre, but not in terms of the centre of gravity of the energy. A possible reason for this is that a model of P-centre location based upon the intensity distribution of a syllable needs to be more complex than that developed by Howell. A weighting function could be incorporated such that events at the onset of the signal are more important than those at the offset. This would bring

the model within the same conceptual framework as the models of musical note perception (Gordon 1987, Vos and Rasch 1981) which are principally concerned with onset events.

The infinite peak clipping experiment is not a good test of either model however. As was discussed in the previous chapter, infinite peak clipping changes a signal to different degrees depending on the amplitude profile of the original signal. Therefore it would be possible that this manipulation would have made more difference to the amplitude profile at the onset rather than the offset (as was the case). In this case the Howell model would have predicted the results, and the data taken as supporting this model. To test the syllabic centre of gravity model, an experimental manipulation which unambiguously affects the amplitude envelope of a signal, and for which specific syllabic centre of gravity predictions can be made.

8.2 A test of Howell's model

An experimental manipulation which would be a better test of the syllabic centre of gravity account was chosen, according to Howell's criteria for testing a determinant of P-centre location.

Howell's criteria for a determinant of P-centre location, which were listed in Chapter 1, are that the determinant should:

> *1) its position in the different stimuli should covary with the subject's adjustments of the stimuli*

If the amplitude profile is a determinant of P-centre location therefore, subject's judgements should vary with a measure of it across stimuli. This was shown in Experiment 3, and not in Experiment 4.

2) its position in the stimulus relative to the physical onset should covary with the subject's adjustments of the stimulus

If the amplitude profile is increased at the offset of a signal, the P-centre set by subjects should shift towards the increase.

3) it should account for perceptual judgments of even timing in both perception and production.

Speakers' produced timing should be affected by an increase of energy at the offset of their utterances.

In this experiment therefore the second criteria was addressed by altering the amplitude envelope of a signal at the offset, without increasing the signal duration. The effect of this manipulation on the P-centre location in perception was addressed. The third criterion was addressed by considering the effect of this variation on speech production. The effect of energy increase at the offset of a stimulus should therefore affect both production and perception, if energy distribution *per se* is a determinant of P-centre location.

The experimental manipulation chosen was partly a replication of Marcus's (1981) finding that increasing the amplitude burst of the final 't' in "eight" does not affect the P-centre location of the word. Increasing the duration of the silent gap before the final 't' burst did however shift the P-centre. The syllabic centre of gravity approach would predict that an increase of energy at the end of the sound (such doubling the amplitude of the final 't' burst) would shift the P-centre back along the sound towards the offset. If this manipulation of the amplitude envelope of the sound had no effect on the P-centre location, the syllabic centre model would have to be modified or rejected as a model of P-centre location.

To amplify Marcus's original experiment and test Howell's third criterion, the effect of instructing speakers to increase the 't' burst at the offset of "eight" on their produced timing was investigated. This would be to examine whether this affects the timing in speech sounds. If, as Howell suggests, speakers produce rhythmic speech sequences with regard to the energy characteristics of their utterances, then this task would result in their producing physically asynchronous sequences. If they produce physically isochronous sequences, it would imply that this account of speakers' timing would need modification.

8.3 P-centre Model Predictions

The syllabic centre of gravity model would predict that this manipulation would result in a shift of the P-centre towards the offset of the signal, as the amount of energy at offset would have doubled.

Marcus's model would predict no shift in P-centre. Vos and Rasch's model would also predict no effect since this model relates perceptual beats to events at the onset of a signal.

8.4 Experiment 5a - perception of rhythmic sequences

8.5 Method

8.5.1 Stimuli

A sample was taken of the spoken word "eight" from an adult male native English speaker. This was the first experimental stimulus (referred to as 'eight').

The 't' burst was identified in the acoustic waveform, edited out, amplified by a factor of two, and re-edited onto the sound, to provide the second

experimental stimulus (referred to as '$eight_{burst}$'). This was then identical to 'eight' in all but the amplitude of the final 't' burst. Figure 8.1 shows oscillograms of the two 'eight' stimuli.

The reference sound (50ms of signal correlated noise) was used as a stimulus in the experiment.

8.5.2 Subjects

Three native English speaking subjects took part in the experiment; all had experience of dynamic rhythm setting tasks.

8.5.3 Design

The independent variable was the amplitude of the final 't' burst. There were two levels: normal amplitude and doubled amplitude.

The experimental procedure followed a full rhythm setting paradigm; all the stimuli were set against each other, the A - A interval was 1250ms.

With three stimuli, each subject performed 9 blocks of 4 trials; these were performed in random order. As before, the final A - B settings were recorded. If Marcus's results were replicated, there would be **no difference** in the A - B interval settings when eight and $eight_{burst}$ were the stimuli. The A - B settings in both cases would be equal to [A - A]*0.5. The syllabic centre of gravity approach would predict a difference in the A - B interval settings. Specifically that the eight-$eight_{burst}$ interval would be smaller than the $eight_{burst}$-eight interval.

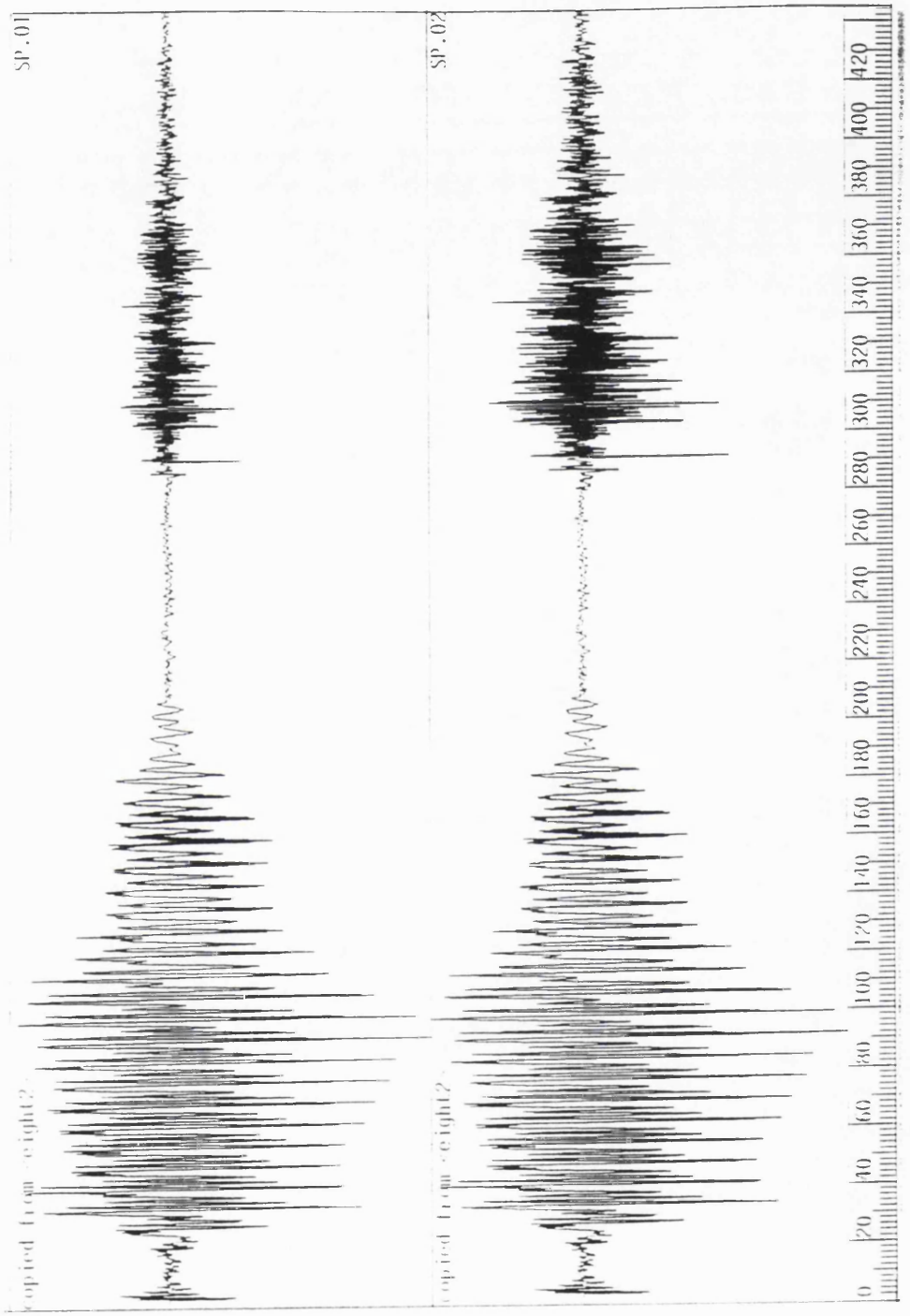


Figure 8.1 Oscillograms of "eight" and "eight$_{burst}$" stimuli.

8.6 Results

Table 8.1 below shows the mean final interval settings for all the stimuli combinations in the dynamic rhythm setting task.

	eight	$eight_{burst}$	ref
eight	622 (24.08)	626 (23.11)	634 (22.85)
$eight_{burst}$	623 (17.21)	615 (22.23)	628 (30.31)
ref	584 (23.48)	615 (34.70)	623 (12.46)

Table 8.1 of means and SDs of final settings(ms) for all stimuli combinations

The means for most of the stimuli combinations are around 625ms (perfect isochrony). This is the size that would be predicted if the subjects were setting the stimuli to physical isochrony. This in turn implies the stimuli have similar P-centres.

The combinations of the reference sound and the normal "eight" appear to deviate from 625ms quite significantly. This indicates that they have differing P-centres.

The combinations of the amplified "eights" and the reference sound do not show such a large deviation from 625ms. This would mean that the two stimuli have more similar P-centres. In the light of the previous paragraph this is difficult to explain; the difference in P-centres should at least be the same for both normal and amplified "eights". The standard deviations around both the means are almost 50% larger for this combination of the stimuli. Some aspect of this combination of stimuli was resulting in a greater variance in final settings. This point will be returned to in the Discussion.

8.7 Analysis

To test the significance of the settings the subjects made, the absolute deviations from isochrony were calculated (the deviation from exact physical isochrony - that is, halfway between A - A). This was calculated by calculating the absolute deviation from [A - A]*0.5 (625ms) for each interval value.

The mean values for each stimuli combination are shown in Table 8.2 below:

	eight	$eight_{burst}$	ref
eight	22.22 (11.07)	17.33 (14.08)	19.88 (13.14)
$eight_{burst}$	12.55 (11.02)	19.11 (10.17)	22.88 (18.45)
ref	40.22 (23.48)	32.33 (11.35)	8.44 (8.98)

Table 8.2 means and SDs of absolute deviations from isochrony for all stimuli combinations (ms)

The standard deviations are not sufficiently heterogeneous for parametric tests - in some cases the means and SDs are very similar. The data was transformed using a square root function. This reduced the means and standard deviations from 21.67ms (16.37) to 4.223 (1.969).

These transformed values were used in a multiple linear regression fitting the equation:

$$y = c + \alpha(x_1)$$

where y=deviations from isochrony, x1=stimuli combination

This was to determine whether the different stimuli combinations were significantly affecting the settings made by subjects, and whether the subjects were a significant source of variation. The two predictors were thus **stimuli combination** and **subject**.

The regression had the equation:

transformed deviations from isochrony = 4.70 - 0.208(stimuli combination)

The predictor **stimuli combination** was significant ($t_{1,78}$ = -2.54, $p < 0.05$).

To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was not significant ($p>0.05$). Subjects were thus consistent with each other in their settings.

Figure 8.2 shows the pattern of deviations from isochrony across the different stimuli combinations.

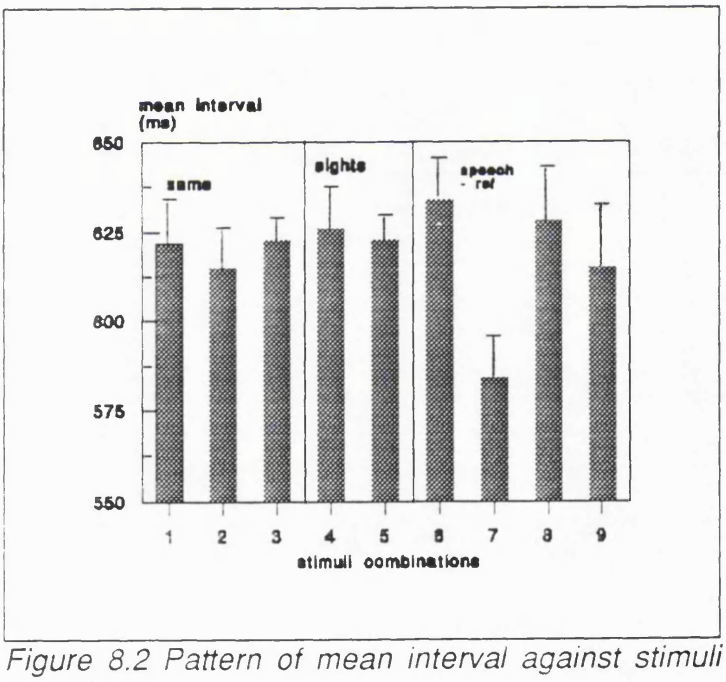


Figure 8.2 Pattern of mean interval against stimuli combinations

Since the stimuli combinations significantly affected the deviations from isochrony in the settings the subjects made, and the subjects were consistent with one another, the interval means (from Table 8.1) were used in the P-centre

determining algorithm. The P-centres calculated were adjusted such that the P-centre of the reference sound was equal to zero. The resultant P-centres are shown in Table 8.3 below:

stimuli	P-centre (ms)
eight	-19.00
$eight_{burst}$	-12.00

Table 8.3 of P-centre settings for the stimuli eight, eight(burst) and the reference sound (ms) (adjusted such that reference sound = 0.0)

The P-centre values are plotted on Figure 8.3. These P-centre values show that the normal "eight" had a P-centre which was later (ie. more negative) than the reference sound. This is congruent with the observations about the mean intervals shown in Table 8.1. This would be predicted since the two stimuli are so very different in duration, spectral content etc..

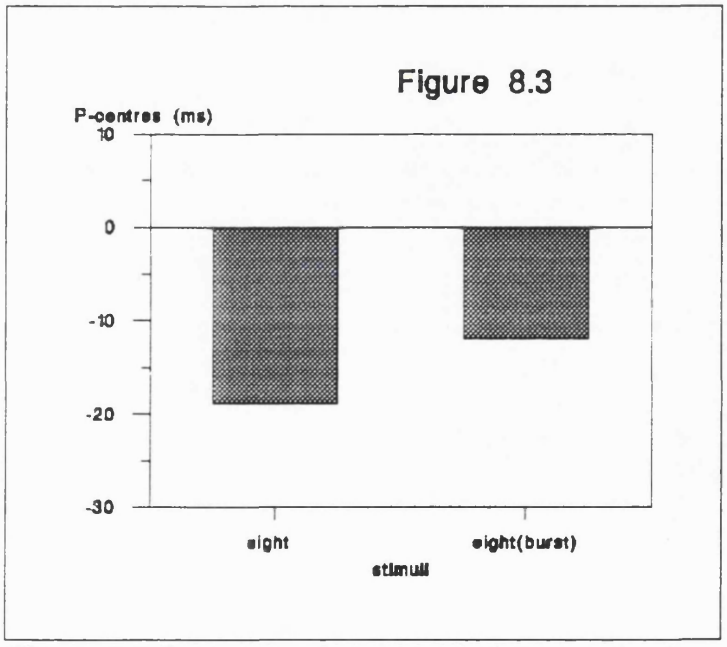


Figure 8.3 P-centres of eight and $eight_{burst}$ stimuli

The P-centres of the eight and the $eight_{burst}$ stimuli are also different. The $eight_{burst}$ has an earlier P-centre (ie. closer to zero) than the normal.

This indicates that there is a very small shift in P-centre caused by the "t" burst amplification. The shift is towards the onset of the signal (since the value is nearer to zero). This is in direct opposition to the syllabic centre of gravity model which predicts a shift towards the offset of the signal (ie. more negative). This P-centre value reflects the intervals set with the eight and $eight_{burst}$ stimuli and the reference sound, shown in Table 8.1, and the discrepancy noted there.

8.8 Howell model predictions

The implementation of Howell's syllabic centre of gravity model described in Experiment 3 was used to provide syllabic centre predictions for the eight stimuli. These are shown in Table 8.4 below.

stimuli	Howell model predictions
eight	-100.5
$eight_{burst}$	-111.4

Table 8.4 Howell syllabic centre of gravity model predictions for P-centre locations of eight stimuli (ms)

The Howell model thus predicts a shift in P-centre of around 10ms away from the onset of the sound when the 't' burst is amplified. This is not consistent with the experimental results, which show a shift of 7ms towards the onset of the stimuli caused by the increase in the 't' burst.

8.9 Discussion

The amplification of the final 't' burst in the word "eight" affected the syllables' P-centre, as measured in a rhythm setting perception experiment. The observed shift in P-centre which occurred as a result of this manipulation was towards the onset of the syllable. The shift in P-centre is not thus in the direction predicted by Howell (1984). His model predicts a shift in P-centre towards the offset of the signal, as the energy at that end increases. This result is also not congruous with Marcus's original experiment, which found that this manipulation had no effect on P-centre location.

What is the reason for this result? Subjective accounts of the experiment from an original pilot study indicated that the amplified "t" bursts started to separate out from the perceptual stream, and were heard as discrete events. Attempts to reduce this effect by slowing down the tempo of the presentation were successful. The amplified "t" burst did still tend to 'syllabify' into a separate event, and subjects may have found that this provided an alternative cue for a rhythm judgement than the P-centre of the entire "eight" syllable; the subjects may have been using these bursts as the rhythmic centre cue for completing the task. This would explain why in both combinations of the amplified "eight" with the reference sound, there is such a large standard deviation. The normal "eight" "t" bursts were not subjectively so perceptually salient. This stimulus was set against the reference sound with a reasonable standard deviation and to a shift from isochrony that would be normally expected.

This P-centre difference between the normal "eight" and the reference sound, provides evidence against the possible criticism that no effect of "t" burst amplification was found because the subjects simply could not do the experiment. This is also supported by the regression of the transformed deviations from isochrony; there was a significant effect of the stimuli

combinations, and not of subjects as a factor. This suggests that the subjects were making consistent settings according to the stimuli, and were not guessing randomly, due to a lack of motivation or inability to perform the task. The significant effect of the stimuli on the regression was due to the trials where the reference sound was a stimulus. This is shown in Table 8.1; the intervals display large deviations from isochrony shown when the two stimuli are the speech sounds, suggesting that their P-centres are not different.

Thus, in this experiment, Marcus's original finding that increasing the amplitude of a syllable final "t" burst does not shift the P-centre of the syllable was partly replicated. This indicates that a syllables' P-centre is not influenced by the amplitude variation over the entire signal as Howell (1988) predicts.

8.10 Experiment 5b - the production of rhythmic sequences

This was not strictly an experiment; instead speakers were instructed to produce rhythmic sequences containing the word "eight" uttered in different manners. The onset to onset intervals of their even sequences were measured in order to detect any anisochronies in these perceptually regular sequences. Such anisochronies would indicate that the different "eights" had different P-centres.

8.11 Method

Rhythmic speech from two speakers was taken. Both speakers were adult native English speakers, one male (SHS) and one female (SKS). The speakers were instructed to produce the word eight rhythmically, until they were satisfied that they were producing rhythmic sequences. They were then requested to produce similar even sequences, with the "t" bursts of the "eights" amplified

(obviously, the speakers could not produce perfectly doubled bursts, but any increase that was regular was accepted). No explicit instruction as to tempo was given - the speakers were told to produce the even sequences at the most comfortable rate.

These sequences provided data about each speaker's timing in normal eight and $eight_{burst}$ sequences. The speakers were also instructed to produce sequences of "eights" with alternate "t" bursts increased in amplitude. There were several practice trials, when the speakers and the experimenters were satisfied with the rhythmicity and the "t" bursts, the sequences were recorded. The recording of the sequences was performed in blocks of sequence type to enable practice trials. The order was randomized for each subject. The recordings were carried out in a sound treated room.

The sequences were digitized at 20kHz. The beginnings and endings of the recorded sequences were not digitized, as these were often less regular (due to phenomena such as phrase final lengthening). As described in the procedure for the production section of Experiment 2, only sequences which were acceptable were digitized - this resulted in uneven numbers of speech items in each sequence type for each subject (see Results). The physical onsets of the utterances were annotated by hand, again using the procedure described in Experiment 2. The onset to onset intervals were measured. Any deviation of the timing caused by the production of alternate eight-$eight_{burst}$ sequences would be shown as a difference in the onset - onset intervals. Thus if the P-centres of the speaker's eight were affected by producing $eight_{burst}$, there would be a difference between the onset to onset duration for the eight to $eight_{burst}$ intervals compared to the duration of the $eight_{burst}$ - eight intervals.

8.12 Results

The Table 8.4 below shows the means and SD's of the onset to onset intervals of the utterances in the different rhythmic sequences, for each speaker. The alternating sequence intervals are split into two sets; the eight to $eight_{burst}$ intervals and the $eight_{burst}$ to eight intervals. This is to show any differences in the size of these intervals.

Figure 8.5 shows these patterns of inter-onset production intervals for the different sequences, for each subject.

			alternating sequences (ms)	
subjects	eight alone (ms)	$eight_{burst}$ alone (ms)	eight - $eight_{burst}$	$eight_{burst}$ - eight
SKS	706.5 (28.00) n = 13	756.2 (19.3) n = 13	882.7 (28.6) n = 22	890.3 (29.0) n = 21
SHS	800.9 (33.01) n = 7	856.6 (19.99) n = 8	1037.4 (41.4) n = 6	1023.3 (34.2) n = 8

Table 8.5 of onset to onset intervals of eight alone, eight(burst) alone, and alternating eight-eight(burst) sequences for both speakers SKS and SHS (means and SD's and no. of stimuli)

A clear trend can be seen across the three types of sequences, for both speakers. The onset to onset intervals become longer over the eight alone, $eight_{burst}$ alone, and alternating sequences. This means that the speed of the sequences slowed over these sequences - both speakers were producing the sequences at a reduced rate. This was not due to fatigue since the sequences were not collected in this order. Both the speakers felt that sequences containing the $eight_{burst}$ utterances were difficult to produce rhythmically, and the

reduction in speed probably represents attempts by both speakers to produce these sequences evenly.

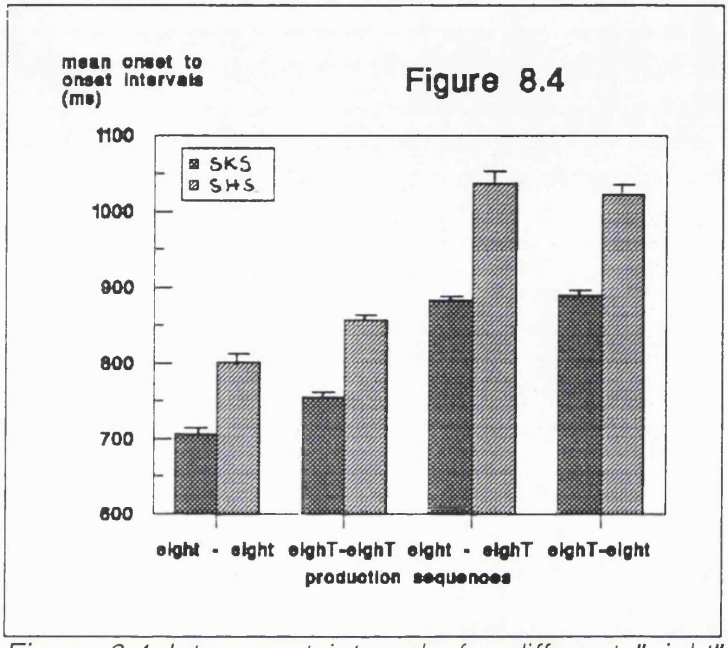


Figure 8.4 Inter-onset intervals for different "eight" spoken sequences, for each speaker

For the alternating sequences, there is no apparent difference in the onset to onset intervals for eight to $eight_{burst}$, compared to the $eight_{burst}$ to eight intervals, for either speakers. The standard deviations for these alternating sequences are of the same order as those for the eight and $eight_{burst}$ alone sequences, for each speaker. This indicates that the alternating sequences were not merely randomly produced by each speaker.

The difference between the intervals for the alternating sequences was tested statistically for each subject using unrelated t tests. For each speaker therefore the t test was used to compare the mean of the eight - $eight_{burst}$ onset to onset intervals against the mean of the $eight_{burst}$ - eight onset to onset intervals. For both speakers the t tests were not significant (for SKS $t_{40} = -0.85$, $p = 0.40$, for

SHS $t_{11} = 0.68$, $p = 0.52$). This meant that there was no significant difference between the onset to onset intervals of the eight-eight$_{burst}$-eight sequences. This implies that the P-centres of the speakers 'eights' are not affected by the increase in amplitude at the syllable offset.

8.13 Discussion

Thus the eight to eight$_{burst}$ sequences were produced by both subjects with even intervals between each utterance, as in the eight to eight and eight$_{burst}$ to eight$_{burst}$ sequences. Instructing subjects to produce alternately increased "t" bursts in perceptually isochronous sequences does not cause them to produce physically anisochronous sequences.

Anisochrony would be expected if the "t" burst amplification affected the production P-centre of the syllable. Therefore it must be concluded that producing increased "t" bursts at the offsets of syllables does not affect the P-centre. This thus supports Marcus's original finding that the P-centres of syllables in perception are not affected by amplitude changes at the offset, and extends the finding to speech production.

8.14 Overall Discussion

The first experiment showed that the doubling of a syllable final "t" burst did not shift the P-centre of that signal towards the offset, as predicted by Howell's syllabic centre of gravity model of P-centre location. The results were not easy to interpret; the amplification of the "t" burst appeared to shift the P-centre towards the onset of the signal. No model of P-centres would predict this result (which is very small - 7ms), and it is certainly counterintuitive. This experiment does not completely replicate Marcus's original finding.

When the amplified eight is set against the reference sound, the final intervals set indicate that the amplified "t" interfered with the settings, and was sometimes used as the cue for rhythm setting. This possibility of alternative cues was tested by Whalen et al (see Method Chapter) and rejected. However, as mentioned earlier their methodology was unsure, and they also did not use manipulated speech as in this experiment, which may well provide distracting cues. Seton's (1989) work described in Chapter 4 indicted that such an effect might well occur.

It is worth noting that Marcus did not use a reference sound in his amplified "t" burst experiments. Therefore any problems that subjects might have setting the reference sound with stimuli that have been manipulated in this way would not have been noted before.

The results of the production experiment support the suggestion that offset "t" burst amplification does not affect P-centre location. Two speakers were capable of producing normal eight/eight sequences that were perceptually isochronous. They could also produce such sequences with increased "t" bursts that sounded evenly timed. Finally, they could produce perceptually even "eight/eight$_{burst}$" sequences; these sequences were produced with physically isochronous intervals - the production of alternate "t" bursts was not affecting their timing. In production, as in perception therefore, increases in intensity at the offset of the syllable does not affect the P-centre of the signal.

This production result is not predicted by Howell's 1984 model. He described an experiment in which speakers produced perceptually isochronous sequences of vowels that varied in duration. They altered their timing to account for the duration of the sound they produced, and Howell concluded that they were making these adjustments according to knowledge of the intensity

characteristics of the syllable they were producing. This experiment was replicated by Fox and Lehiste (1987).

The present experiment is not very similar to Howell's original finding, however. He instructed speakers to vary the duration of the vowels. Speakers in this experiment were varying the loudness of a syllable final segment. A possible hypothesis to explain the observed lack of difference caused by "t" burst amplification is that speakers do not vary their timing according to the entire amplitude distribution *per se* of their utterances; they may vary their timing due to the duration of the produced syllables. A second hypothesis is that they vary their timing according to the intensity distribution of their vowels, but not the consonants.

Overall this experiment generally replicates Marcus's original finding; increasing the amplitude of a syllable final "t" burst in produced speech does not affect the pattern of onset to onset intervals in way which indicates P-centre shifts; doubling the amplitude of a syllable final "t" burst has ambiguous effects in perception experiments, that may reflect a P-centre shift towards the onset, or is an experimental artifact arising due to alternative perceptual rhythmic cues.

Hypotheses can be made about parameters which affect P-centre location, based upon these results:

1) Intensity events at the offset of a signal are less important than those at the onset in regard to their affect on the P-centre of the signal.

2) There are differences in the spectral characteristics of a voiceless plosive "t" (as was amplified in this experiment) and a vowel (as used by Howell). These are reflected in the perceptual salience (known as sonority) of the segments. Perhaps there are

frequency dependant aspects to the effect of intensity on P-centre location; intensity changes within certain frequency bandwidths may be more important than those within others.

3) Both of the above possibilities may occur together - they are not mutually exclusive.

These possibilities will be tested in the next experiment, in order to develop a new model of P-centre location that can account for the results described so far.

Chapter Nine
Experiment six
The effect of stimulus rise time on P-centre location

Abstract
The local models of P-centre location which have appeared in the musical beat literature have manipulated the rise times of signal. In this experiment the rise time of speech signals was manipulated. An affricate initial syllable, a semi-vowel initial syllable and a vowel wās examined. This was to test whether ramping the onsets of speech sounds affects the P-centres of the speech sounds, and also whether the P-centres of different speech sounds are affected differentially by this manipulation.

9.1 Introduction

Marcus's model of P-centre location determines P-centre location by considering the durations of different portions of the stimulus. His choice of these parameters was principally a consequence of his use of these as variables in his experiments. Stimulus rise time was the variable manipulated by Howell (1984) when he demonstrated that P-centres vary with the amplitude envelope of a signal in both speech and non-speech. Howell's experimental paradigm did not enable the calculation of P-centres from the intervals set (see Chapter 1). There was no fixed A - A interval, meaning the overall tempo of the sequences was not held constant. However he reported a significant effect of the stimulus rise time on the intervals set by the subjects.

Other workers in the field of non speech sounds had shown that stimulus rise time has a significant effect on P-centre location (or perceived attack time, or perceived onset time, both of which are defined as P-centres are in this thesis).

The results of Vos and Rasch (1981) indicated that manipulating the rise time of a signal altered its perceived onset; Gordon (1987) in a descriptive account, found that the perceived attack time of musical tones varied with the rise time of the tones.

Although they would seek to define P-centres in very different terms, the experimental manipulations of Whalen Cooper and Fowler generally follow similar patterns to those of Marcus (in terms of duration variation) and Howell (in terms of amplitude manipulations). In their 1986 paper they demonstrated that if successive portions of frication are excised from a "sha" sound, the resulting stimuli form a continuum between "sha" "cha" and "ta". They showed that subjects perceived these differences categorically. The P-centres of these stimuli were affected in a linear manner, regardless of the phonetic category boundaries. They concluded that P-centres are not affected by phonetic identity.

After similar experiments where the stimulus continuum "sa" - "sta" was created (by the insertion of different duration gaps) and where either frication duration or vowel duration was held constant, they made the more controversial claim that the stimulus rise time had no effect on P-centre location. Controversial because this was a parameter which they had not manipulated experimentally (as described in Chapter 1).

The results of the infinite peak clipping experiment described in Experiment 5 indicated that while this manipulation of the amplitude envelope affected the P-centres of the speech signals, it was the change at the onset of the signal that was leading to the shift in P-centre towards the onset. The results of Experiment 5 showed that, in both production and perception, alteration of the amplitude envelope at the offset did not generally affect the P-centre of the signal.

This is all evidence that onset amplitude events determine P-centre location, as opposed to amplitude characteristics over the entire signal. A method of describing onset amplitude events is to describe the *rise time* of a signal; this is the parameter varied by Vos and Rasch (1981) and Howell (1984), and measured by Gordon (1987). It is defined acoustically as the duration of the period during which the amplitude envelope passes between 10% to 90% of peak amplitude. The shorter the rise time of a signal, the faster the signal reaches peak amplitude (in musical terms, this is a fast 'attack'). The longer the rise time, the more gradual the increase in amplitude at the onset of the sound. This parameter of rise time, is that which Whalen et al (1986) claimed specifically did not affect the P-centre of a speech sound.

9.2 Experimental aims

In this experiment the rise time of speech signals will be manipulated to determine the effect of this manipulation upon P-centre location. This manipulation will be performed upon different types of speech items, to determine whether the strength of the effect is equal across stimuli; that is, whether there is a stimulus dependent attribute of this alteration. Is changing the onset of a vowel equivalent to changing the onset of a fricative? The motivation for this hypothesis is the sonority principle (Ladefoged 1982, Selkirk 1986). Sonority is a term from linguistics which describes the perceptual loudness of a speech sound compared to other sounds - the perceptual strength of a sound. Sonority is a descriptive, linguistic term and hard to define acoustically, and can vary not only across speech sounds but also across speakers. It is however a useful conceptual starting point for differentiating between the perceptual strength of different speech sounds. Table 9.1a below shows a scale of phonetic items, and their associated sonority values (after Ladefoged 1982, p222).

sounds	sonority values	examples
low vowel	6.0	f**a**ther
mid vowel	5.7	h**ea**d
high vowel	5.3	f**ee**t
flaps	5.2	**r**ed
laterals	5.2	**l**ob
nasals	4.9	si**ng**
voiced fricatives	3.8	**z**oo
unvoiced fricatives	1.0	**s**ee
voiced stops	0.2	**d**ump
voiceless stops	0.1	**p**in

Table 9.1a Different speech sounds and associated sonority values - letters in bold are examples (After Ladefoged 1982)

The rise time of speech items were manipulated by *ramping* the onsets; that is, by imposing a scaling linear ramp of different durations on the signal, depending on the rise time required. A distinction must therefore be drawn between the amount by which a signal is *ramped*, and the resultant *rise time* of that signal. The former is quantified by the controlled software manipulation, the latter is an attribute of the signal which must be measured.

9.3 General method

Three experiments were thus run. In each a naturally produced speech sound was taken, and its onset ramped linearly. The stimuli were then used in dynamic rhythm setting experiments to determine the P-centres. An effect of ramping the stimulus rise time on the P-centre would be shown by a systematic variation in P-centre with degree of ramping. The first experimental hypothesis is that the P-centre will be affected by rise time of the stimulus and will be shifted back in time linearly as ramping increases. The second hypothesis is that there will be a difference between the effect of ramping according to the sonority of the stimuli.

9.4 Model predictions

The approach of Fowler *et al* would predict that the manipulation of the onset rise times of natural speech would not affect the P-centres of the signal since the underlying articulatory information would remain unchanged while the intensity profile is altered. This would apply to all the speech stimuli.

The model of Marcus would predict that none of these manipulations would have an effect since the durations of the syllable portions remain unchanged. This would apply to all the speech stimuli. However his model would predict a change in the P-centres if the ramping affected the defined vowel onset (the peak increment in mid-band spectral energy).

The model of Howell would predict that increasing the amount of ramping - increasing the rise time of the stimulus - would shift the P-centres of all the speech sounds back in time, and that this effect would be the same for the different speech sounds.

The experimentally measured P-centres of the stimuli will be plotted against the measured rise times of the stimuli (measured rise time is used rather than the original amount of ramping since the natural speech sounds are already ramped to some degree). This will show if there is an effect of the stimulus rise time on P-centre location. If the regression lines of these plots have identical gradients, this will show if the effect of ramping is the same for each speech sound, or if there are differences.

9.5 Experiment A. Ramping a CV (fricative - vowel) syllable

A 494ms vowel ("ah"), produced by an adult male native English speaker, was edited together with 210ms of synthesized frication. This produced the syllable "cha", which sounded natural. The frication had an 'immediate' rise time of 0ms. This stimulus was then linearly ramped at the onset, using software, to three different ramping values. These were 10ms, 60ms and 120ms. A synthetic fricative portion was used to ensure that the signal had a steady state and no differences, of spectral or intensity aspects, that might be affected by ramping and lead to spurious results. The maximum amount of ramping (120ms) was constrained by the duration of the "cha" frication, and this was in turn constrained by the psychological plausibility of the syllable as still being speech like. The three stimuli varied perceptually between a "cha" (short rise time) and a "sha" (long rise time). This effect of fricative rise time on phonetic perception is well documented (Gerstman 1957). The duration of the stimuli and the vowel onset time did not vary. This set of stimuli are referred to as the "sha" stimuli.

These three "sha" stimuli were used in a dynamic rhythm setting task, with the reference sound. This led to 16 stimuli combinations. The A - A interval used was 1600ms; this is longer than the usual 1250ms generally used throughout this thesis. The 1250ms tempo could not be used since the stimuli, at 704ms duration, were to long. The 1600ms A - A interval was the most suitable tempo

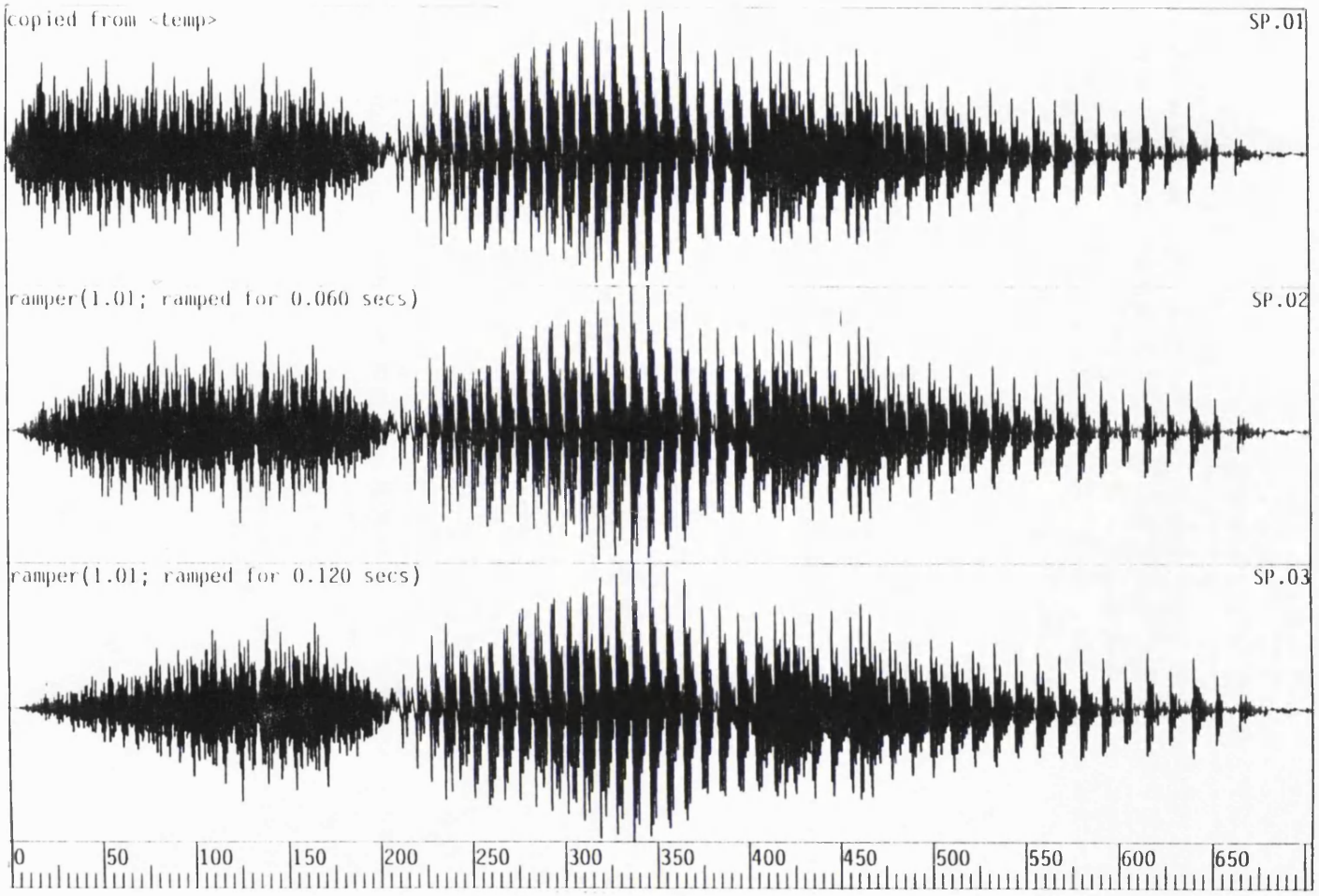


Figure 9.1 Oscillograms of "sha" stimuli - ramped to 10ms, 60ms and 120ms.

as determined by pilot trials, as not being so short that adjustments could not be made, and not so long that the tempo was too slow to make accurate settings.

Figure 9.1 shows the oscillograms of the three "sha" stimuli.

9.6 Experiment B. Ramping a CV (semi-vowel - vowel) syllable

The syllable "wa" was produced by an adult male native English speaker. The duration of the consonant /w/ (as defined by Marcus's vowel onset algorithm) was 375ms, the duration of the vowel was 58ms. Since /w/ is a 'liquid' or semi-vowel in phonetic terms, however, the consonant - vowel distinction is much less clear than in the "sha" stimuli. The broad 'consonant' /w/ portion was ramped using software. Three levels of ramping were 0ms (ie naturally produced), 120ms and 240ms. The speech signal was already ramped to a marked degree at the onset, due to the manner of articulation of /w/ leading to a naturally ramped onset. The actual rise time of the natural "wa" was not therefore equal to zero (see Results section for details of measured rise times of all stimuli).

The amount of experimental ramping was constrained by the duration of the /w/ portion; and the fact that the onset was already ramped naturally. The levels of ramping were therefore increased. The naturally ramped onset of "wa" meant that ramping the onset did not alter the syllable's phonetic structure. The onsets were noticeably 'softer'.

The three "wa" stimuli were used in a dynamic rhythm setting task along with the reference sound. There were thus 16 stimuli combinations. The A - A interval used was 1250ms.
Figure 9.2 shows the oscillograms of the "wa" stimuli.

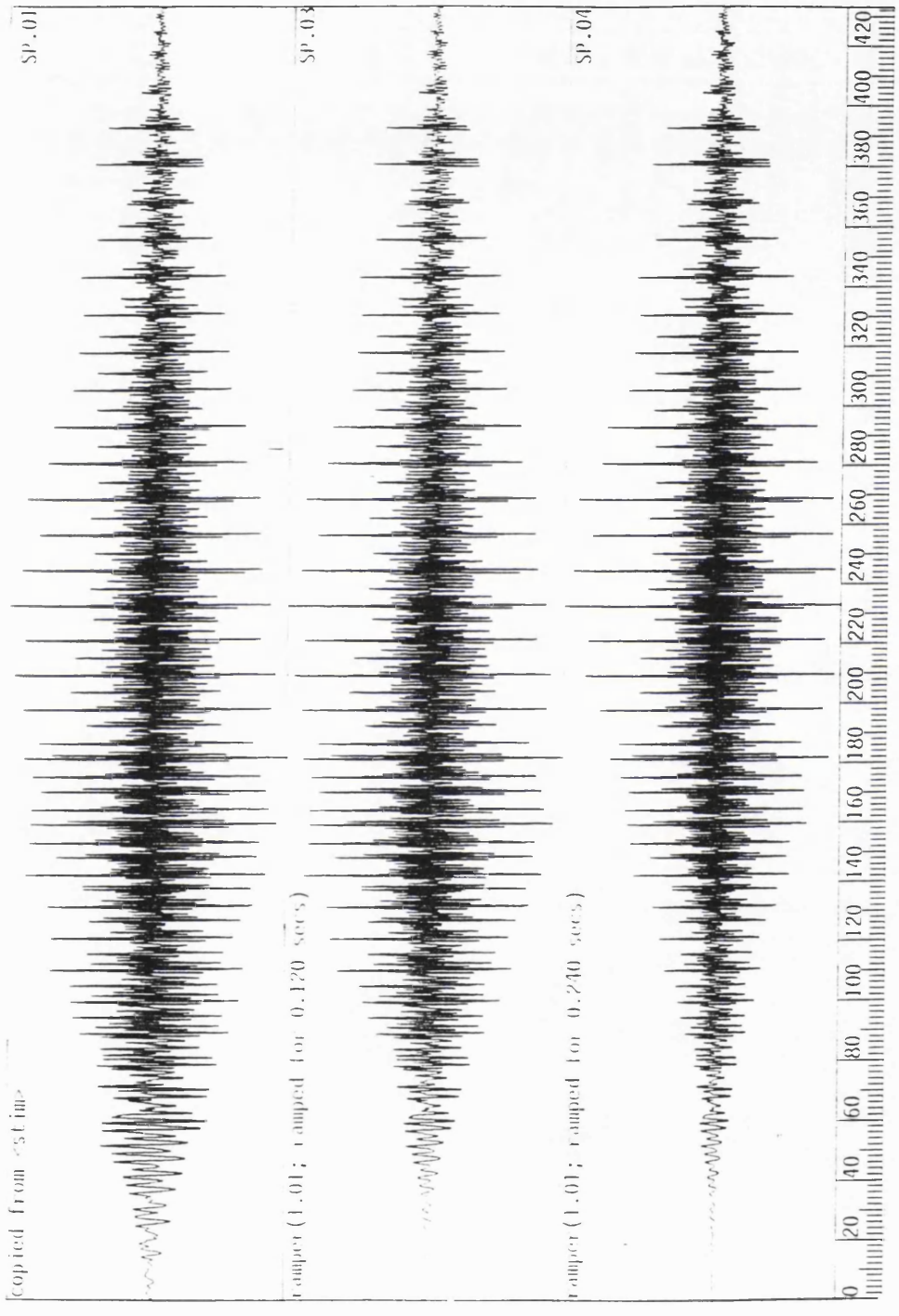


Figure 9.2 Oscillograms of "wa" stimuli - ramped to 0ms (natural onset), 120ms and 240ms.

9.7 Experiment C. Ramping a vowel alone

A vowel /ae/ was produced by an adult female native English speaker. The duration of the vowel was 213ms. The onset of the vowel was naturally fast. The vowel was ramped to three levels; these were constrained by the shorter duration of the vowel and were 10ms, 50ms and 90ms. Perceptually this led to two of the stimuli having a 'softer' onset. The vowel was still speech like and identifiable; its phonetic identity was not degraded.

The three stimuli were used in a dynamic rhythm setting task, with the reference sound. There were therefore 16 stimuli combinations. The A - A interval used was 1250ms.

Figure 9.3 shows the oscillograms of the "ae" stimuli.

9.8 Spectral content

Figure 9.4 shows spectrogram of the first token the "sha" stimulus set. This shows a period of high frequency noise in the fricative "sh" portion which is not present in the onset of the "ae" and "wa" stimuli (see APPENDIX for "wa" spectrogram).

9.9 General Method.

All types of stimuli were used in dynamic rhythm setting tasks. The value of the A - A interval was 1250ms (the same as all other experiments) except for Experiment A. The duration of the sounds was too long for the 1250ms interval to be used; an interval of 1600ms was used instead.

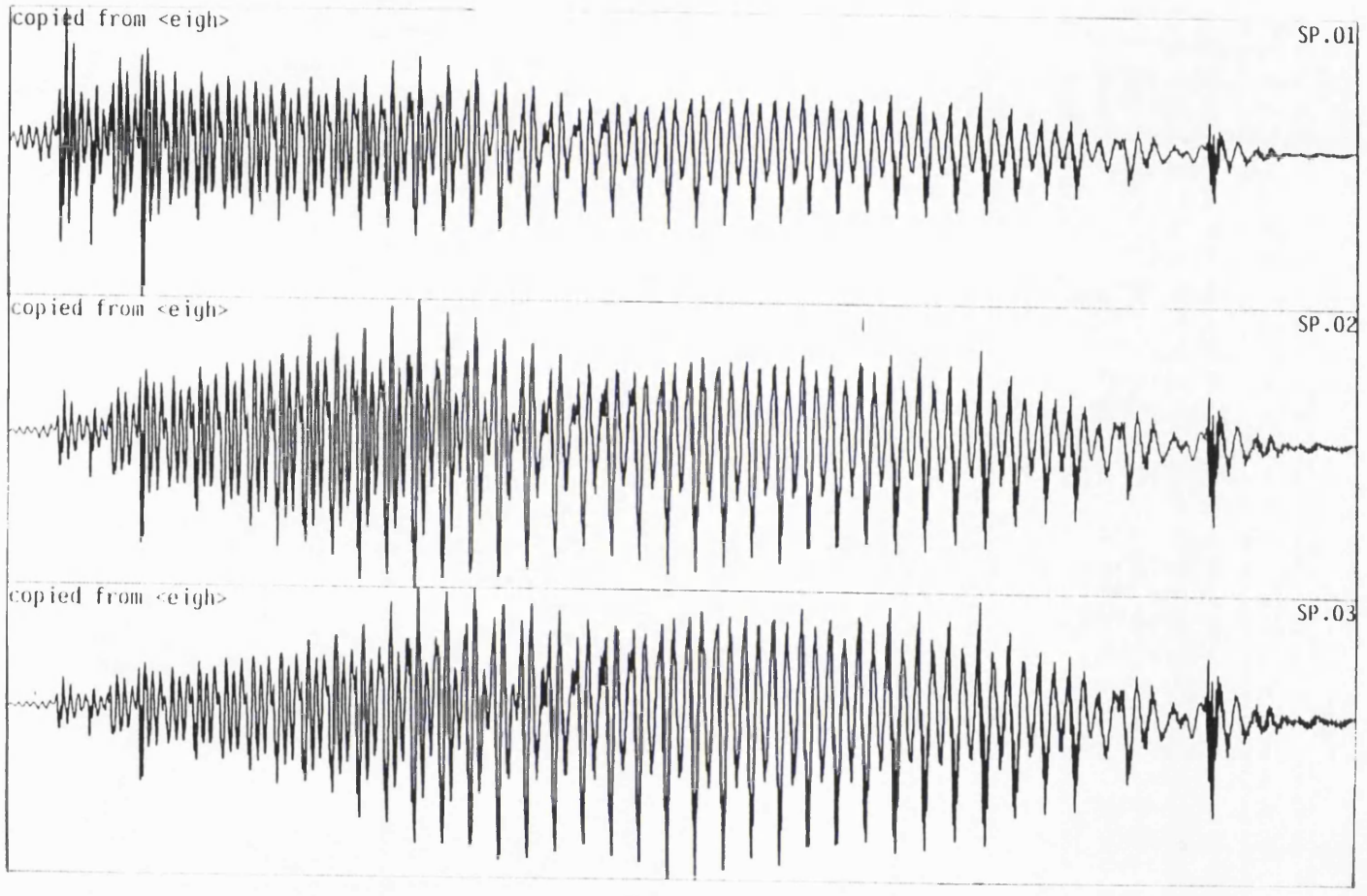


Figure 9.3 Oscillograms of "ae" stimuli - ramped to 10ms 50ms and 90ms.

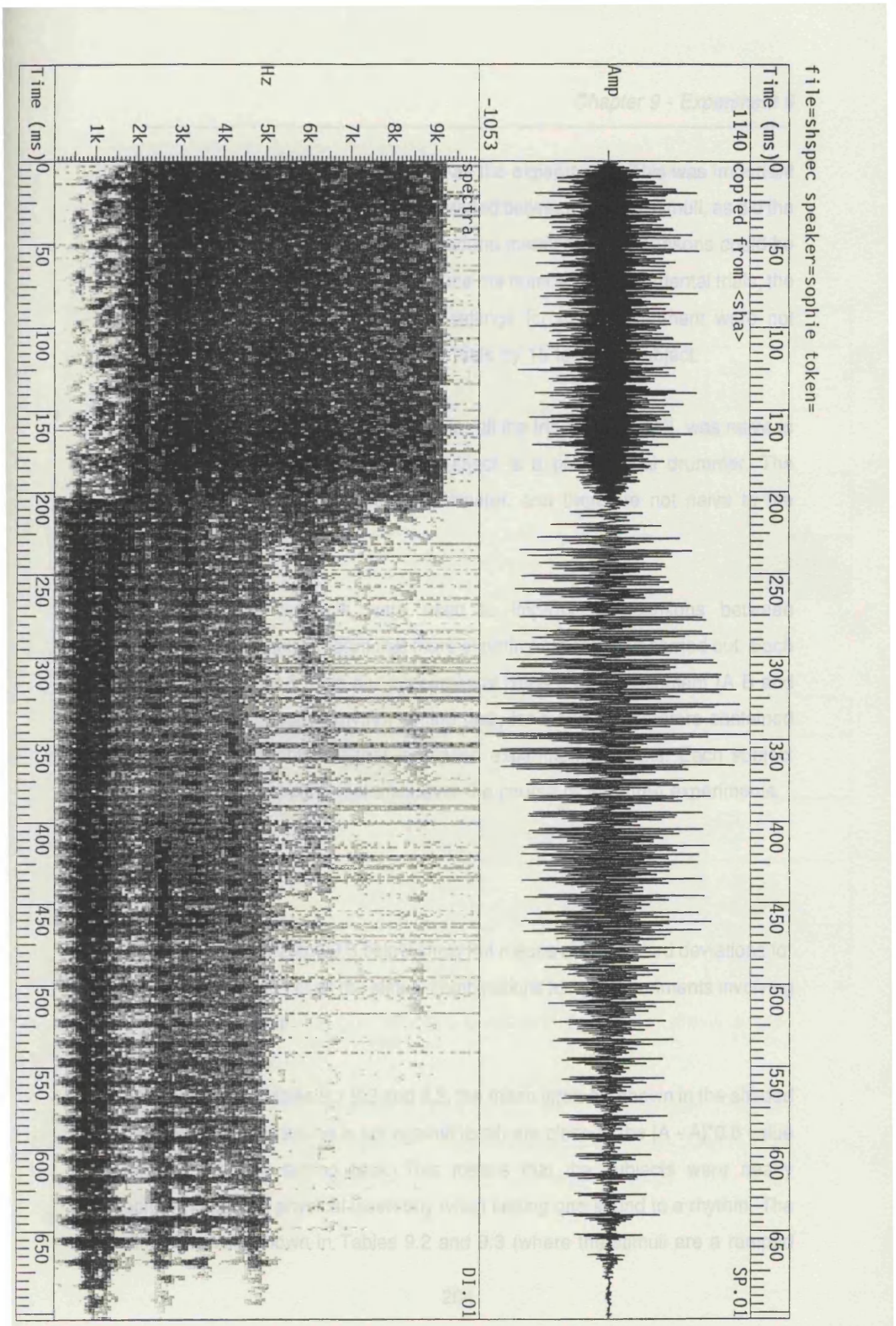


Figure 9.4 Oscillogram and Spectrogram of first "sha" stimulus

The reference sound was included in all the experiments. This was important since the absolute amount of ramping varied between speech stimuli, as did the tempo; the inclusion of the reference sound meant that comparisons could be made across the experiments. To reduce the number of experimental trials, the reference sound - reference sound settings for each experiment were not collected; this reduced the number of trials by 15 for each subject.

Two experienced subjects were used for all the trials. Subject DL was naive to the aims of the experiments; this subject is a professional drummer. The second subject SKS was the experimenter, and therefore not naive to the experimental manipulations.

The same two subjects were used to improve comparisons between experiments, and also meant that more experiments could be carried out. Each subject performed 5 trials per experimental block. Each experiment (A B and C) being a full dynamic rhythm setting task, for n stimuli, therefore contained nXn experimental blocks; that is sixteen experimental blocks. Each subject performed 240 experimental trials over the course of the three experiments.

9.10 Results

The Tables 9.1, 9.2 and 9.3 below show the means and standard deviations for the interval settings of all the stimuli combinations for the experiments involving the ramped stimuli.

For each of the Tables 9.1 9.2 and 9.3, the mean intervals shown in the shaded cells (where the a sound is set against itself) are close to the [A - A]*0.5 value for each rhythm setting task. This means that the subjects were nearly achieving perfect physical isochrony when setting one sound to a rhythm. The other intervals shown in Tables 9.2 and 9.3 (where the stimuli are a ramped

semi vowel and a ramped vowel respectively) vary across the different stimuli combinations, implying that for these combinations the subjects had to make settings that deviated from physical isochrony in order to acheive perceptual isochrony. That is, that these stimuli had different P-centres.

"sha" stimuli	$ramp_{10ms}$	$ramp_{60ms}$	$ramp_{120ms}$	ref
$ramp_{10ms}$	809.64 18.01	794.53 23.27	793.93 23.17	960.80 46.6
$ramp_{60ms}$	801.33 23.93	797.93 24.02	794.87 18.91	940.20 35.5
$ramp_{120ms}$	802.40 26.32	796.93 23.43	800.93 28.34	961.00 45.8
ref	663.00 47.10	630.00 28.2	661.80 42.0	

Table 9.1 of means and SD's for final interval settings for "sha" stimuli ramped at the onsets (ms) (A - A = 1600ms)

"wa"	$ramp_{0ms}$	$ramp_{120ms}$	$ramp_{240ms}$	ref
$ramp_{0ms}$	628.07 18.36	602.82 18.79	592.83 19.21	665.3 32.2
$ramp_{120ms}$	645.64 12.87	635.90 19.01	610.08 16.10	690.30 29.16
$ramp_{240ms}$	644.88 17.39	629.15	622.23 13.63	697.50 14.48
ref	584.00 20.36	551.90 24.25	555.10 23.13	

Table 9.2 of Means and SD's for final interval settings for "wa" stimuli ramped at the onsets (ms) (A - A = 1250ms)

For the intervals shown in Table 9.1 (the ramped fricative stimuli) all the values are close to [A - A]*0.5 (800ms). This indicates that for all these stimuli perceptual isochrony could be acheived with near physical isochronous

intervals, and thus that the P-centres of these stimuli do not vary much. These observations will be examined in the following sections.

"ae"	$ramp_{10ms}$	$ramp_{50ms}$	$ramp_{90ms}$	ref
$ramp_{10ms}$	620.75 15.66	608.14 6.77	599.67 7.98	620.30 14.73
$ramp_{50ms}$	627.91 13.90	615.75 5.39	612.38 8.81	639.00 19.88
$ramp_{90ms}$	637.40 8.17	628.78 5.93	623.80 17.96	652.80 26.72
ref	619.67 9.86	604.00 19.06	600.90 17.08	

Table 9.3 of Means and SD's for final interval settings for "ae" stimuli with ramped onsets (ms) (A - A = 1250ms)

9.11 Analysis

The absolute deviations from isochrony were calculated for all the sets of intervals (that is the value of each interval minus half the A - A interval value). This gives a measure of the magnitude of the deviation from isochrony needed to successfully adjust the signals to a perceptually isochronous rhythm.

sha	1	2	3	ref
1	16.79 (11.00)	20.40 (11.29)	17.53 (15.71)	171.44 (40.82)
2	17.60 (15.57)	19.27 (13.76)	16.73 (9.25)	189.14 (45.34)
3	20.53 (15.72)	19.60 (12.13)	19.64 (19.71)	231.75 (23.32)
ref	165.83 (58.05)	189.14 (45.34)	151.56 (39.93)	

Table 9.4 means and SD's of absolute deviations from isochrony of intervals set with "sha" stimuli (A - A = 1600ms) (ms)

"wa"	1	2	3	ref
1	17.50 (10.09)	23.636 (16.72)	36.17 (21.48)	41.50 (30.46)
2	21.58 (13.74)	17.50 (12.37)	16.58 (14.20)	65.30 (29.16)
3	22.75 (17.88)	17.58 (15.34)	19.64 (19.71)	72.50 (19.48)
ref	41.00 (20.35)	73.10 (24.24)	69.90 (23.73)	

Table 9.5 means and SD's of absolute deviations from isochrony of intervals set with "wa" stimuli (A - A = 1250ms)(ms)

The means and standard deviations of these deviations from isochrony are shown for each set of stimuli in Tables 9.4, 9.5 and 9.6.

"ae"	1	2	3	ref
1	13.75 (7.06)	16.85 (6.76)	25.33 (7.98)	10.10 (11.31)
2	10.90 (8.95)	9.25 (5.39)	12.87 (8.39)	18.60 (15.12)
3	12.40 (8.16)	6.44 (2.12)	14.00 (10.30)	27.60 (26.72)
ref	9.77 (4.65)	21.40 (18.55)	24.10 (17.07)	

Table 9.6 means and SD's of absolute deviations from isochrony of intervals set with "ae" stimuli (A - A = 1250ms) (ms)

The standard deviations of these mean deviations from isochrony are not heterogeneous - the SDs being in some cases larger than the mean. The data was therefore transformed using a square root function. This reduced the overall means and SD's to 6.341 (4.45) for the "sha" stimuli, 5.41 (2.49) for the "wa" stimuli and 3.603 (1.662) for the "ae" stimuli.

To test the significance of the stimuli combinations upon the deviations from isochrony of the intervals that the subjects set, and for any difference between the two subjects, the transformed absolute deviations from isochrony were regressed against the stimuli combinations and the subjects in a multiple regression fitting the equation:

$$y = c + \alpha(x_1)$$

where y=deviations from isochrony, x1=stimuli combination

This was done for a concatenation of all three sets of stimuli. The predictors were thus **stimuli combination**

The regression has the equation:

transformed deviations from isochrony = 3.63 + 0.226(stimuli combination)

The predictor **stimuli combination** was significant ($t_{1,480} = 15.97$, $p < 0.05$).

To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was significant ($t_{1,480} = 4.25, -2.06$, $p < 0.05$). The two subjects therefore differed significantly in the settings they made.

If ramping the onsets of the stimuli leads to alterations of the P-centres of the stimuli, the effect of different stimuli combinations upon the subjects' settings would be predicted. Therefore the stimuli combinations should account significantly for some of the observed variance.

The subject difference is however a problem. Cooper, Whalen and Fowler (1986, 1988) reported significant differences between subjects when the

durations of attributes of speech stimuli were varied; the subjects' responses were not equally affected by these manipulations.

To examine the differences between the two subjects, the individual mean deviations from isochrony for each stimuli combination were plotted for each subject, for each set of data. Any differences between the subject's mean settings would be shown as differences between the plots.

Figures 9.5 9.6 and 9.7 show each subjects' mean deviation from isochrony for each stimuli combination for the "sha", "wa" and "ae" stimuli respectively. The stimuli combinations are marked 1-15 on each x-axis. These relate to the stimuli combinations as follows:

combinations 1 - 3: trials where the same sound is set against itself

combinations 4 - 9: trials where speech stimuli with differently ramped onsets are set against each other.

combinations 10 - 15: trials where stimuli with differently ramped onsets are set against the reference sound (50ms signal correlated noise, ramped onset and offset)

On each of the figures it can be seen that the subjects make reasonably similar deviations from isochrony on all the trials up to trial 10. Some difference between the subjects' setting would be expected since there will be a degree of constant variation of attention, motivation etc. This is why several trials are performed by each subject. In addition Marcus (1981) in his algorithm for

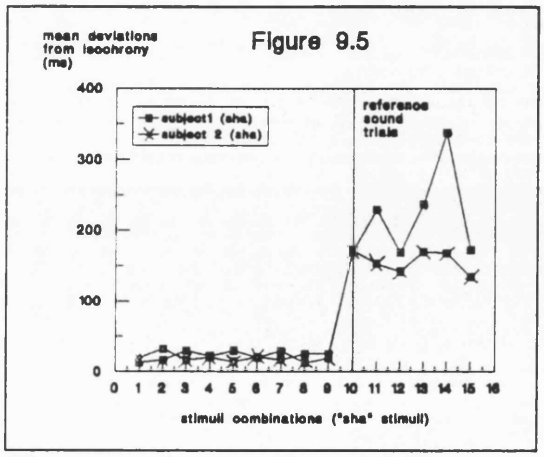


Figure 9.5 each subject's mean deviation from isochrony for each stimuli combination for "sha" stimuli

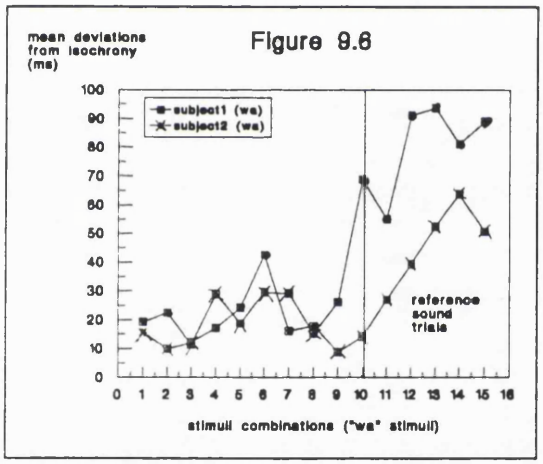


Figure 9.6 each subject's mean deviation from isochrony for each stimuli combination for "wa" stimuli

establishing P-centre values noted that subjects have tendencies to 'overturn' the knob when making their settings; his algorithm compensates for this by utilising intervals in each order (both sound A adjusted relative to B, and B adjusted relative to A). This 'overturning' would not be compensated for in the representation of the intervals shown in these figures, since the values used are absolute (no minus values). Some of the variance must be due to this difference therefore.

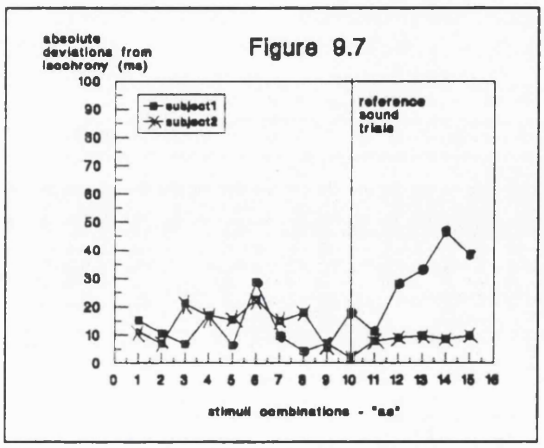


Figure 9.7 each subject's mean deviation from isochrony for each stimuli combination for "ae" stimuli

The principle difference between the two subjects is shown by a deviation of the plots from trial 10 onwards. Thus the subjects behave differently on the trials where the speech stimuli are set against the reference sound.

This raises the concern that this difference reflects the use of a non naive subject (subject 2), who may make more exaggerated settings, albeit unconsciously. This would be a strong reason for not using this subject at all. However on examining all three figures it can be seen that on the reference sound trials, it is subject 1, the naive subject, who is making the larger deviations in their settings.

This result must therefore be explained. There are two alternative suggestions. The first is that, contrary to general opinion about P-centres, there are differences between listeners in the physical deviations that lead to perceptual regularity. This could be due to production differences, or hearing differences, for example.

A second explanation is that there are differences between the subjects' rhythm setting patterns that are affected by attributes of the task other than the P-centres of the stimuli. On the trials 1 - 9, for all the stimuli types, the subjects make broadly similar deviations in their settings. This argues against their perceptions of the P-centres differentially affecting their settings. Instead they are different only on the reference sound trials, and on these trials subject 1 makes very large deviations. A tentative hypothesis is that subject 1's performance on these trials varied much more as a factor of his being a professional drummer. The reference sound trials had an overt percussive element; this subject may have thus been making settings more in accordance with a syncopated rhythmic structure such as might be more normally played, as opposed to the plain marching cadence which is demanded in the dynamic rhythm setting task.

This is conjecture; this hypothesis is backed up by the subjects' similar settings on the speech stimuli only trials, for all three stimuli types. Their intervals were thus treated together to determine P-centres for the stimuli. This was acceptable because the P-centre algorithm makes a P-centre fit based upon all the mean intervals, thus discrepancies in the reference sound trials would be reduced in their effect. The alternative of leaving out the reference trials would not enable any kind of comparison between the sets of P-centres; the alternative of calculating separate P-centres for each subject was not justified by any consistent difference between the subjects over all the trials.

9.12 P-centre calculations

The P-centres were not investigated in terms of the amount by which the stimuli were ramped experimentally. This was because a linear ramp applied to a naturally ramped onset does not necessarily result in a linear onset of the same magnitude of as the ramp. This can be seen clearly in the oscillograms of the

stimuli; the "sha" stimulus has a clear, short original rise time and a constant amplitude over the frication period. The onset rise times are thus similar to the applied ramps. For the "wa" and the "ae" the linear ramp does not result in a controlled variation of onset rise time due to the natural ramping of the onset and the varying amplitude of the signal over time.

A regression of P-centres against the amount of ramping would thus not be an accurate measure of how P-centres vary with the physical onset characteristics of the stimuli. Instead the rise times of the stimuli were calculated. This was done with software, which smoothed the signal, found the peak amplitude, and then established when the smoothed signal passed between 10% and 90% of the maximum amplitude. The onset of the stimuli that the subjects adjusted in the experiment was determined according to the structure of the SFS file, not the onset of the signal. This was the 'physical onset' which is used in the P-centre algorithm, and might not relate to any onset measure of the signal. Therefore the time from the start of the file structure to the 90% or peak amplitude was used as a measure of the rise time of the signal. These values are shown, along with the amount of experimental ramping, in the tables below. The mean intervals set for all the stimuli combinations (as shown in Tables 9.1, 9.2 and 9.3) were used to calculate P-centres for the stimuli. These were then adjusted such that the reference sound P-centres were equal to zero. The resultant P-centres are shown in the Tables 9.7, 9.8 and 9.9 below.

		"sha" stimuli
ramping (ms)	rise time (ms)	P-centre (ms)
0.0	17.0	-148.5
60.0	54.0	-152.5
120.0	110.0	-152.0

Table 9.7 amount of ramping, measured rise time and calculated P-centres for "sha" stimuli

The P-centres of the stimuli are not varying to a great degree with the rise time of the stimulus. This relationship is plotted on Figure 9.8 below.

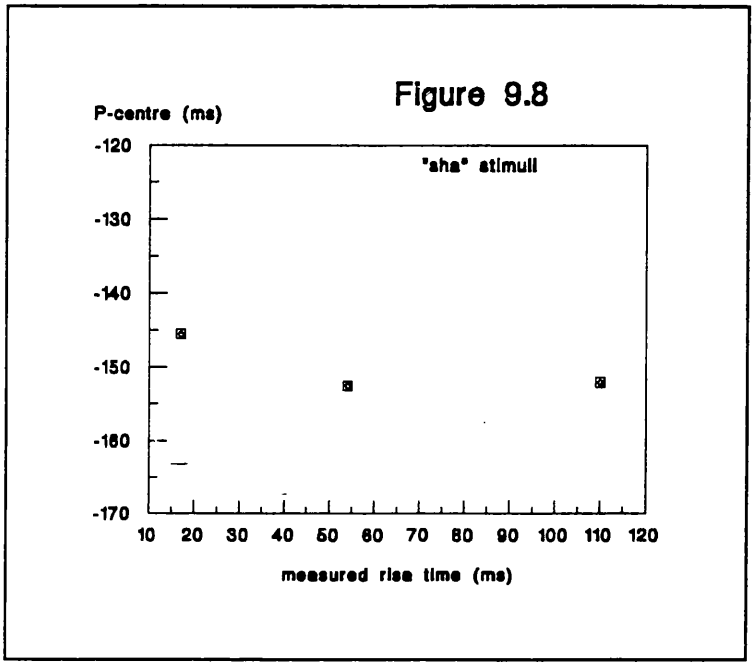


Figure 9.8 P-centres of "sha" stimuli plotted against stimulus rise time

Figure 9.8 shows that there is slight trend of the P-centre and the rise time; the longer rise times are leading to later P-centres. This effect is small.

		"wa" stimuli
ramping (ms)	rise time (ms)	P-centre (ms)
0.0	138.0	-43.25
120.0	149.0	-65.25
240.0	200.0	-71.5

Table 9.8 amount of ramping, measured rise time and calculated P-centres for "wa" stimuli

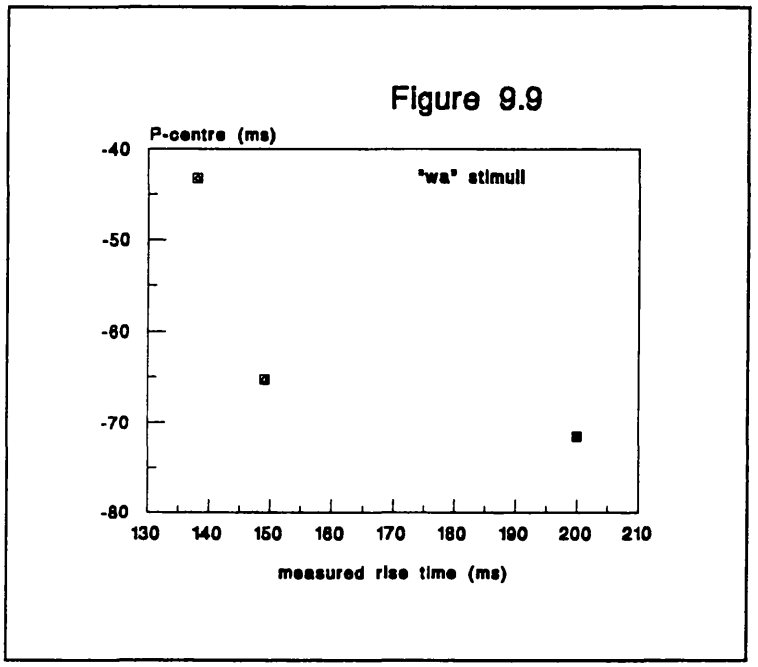


Figure 9.9 P-centres of "wa" stimuli plotted against stimulus rise time

The P-centres of the "wa" stimuli are varying with the rise times of the stimuli. The longer rise times appear to lead to later P-centres.

This relationship is plotted on Figure 9.9. There is a strong effect of stimulus rise time upon P-centre location.

		"ae" stimuli
ramping(ms)	rise time (ms)	P-centre (ms)
10.0	12.0	-3.75
50.0	56.0	-15.25
90.0	64.0	-24.00

Table 9.9 amount of ramping, measured rise times and calculated P-centres for "ae" stimuli

As with the "wa" stimuli, the P-centres of the "ae" stimuli appear to vary with the rise times of the stimuli. This relationship is plotted on Figure 9.10.

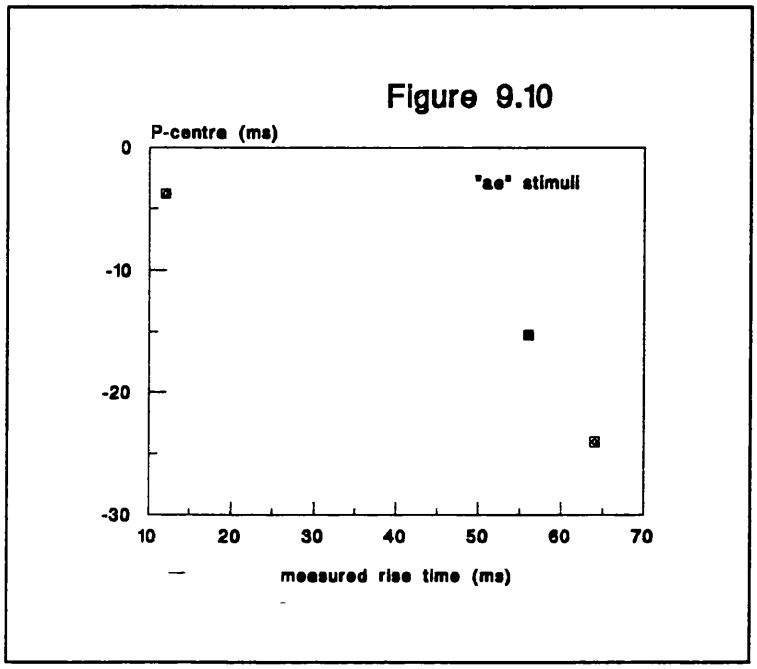


Figure 9.10 P-centres of "ae" stimuli plotted against stimulus rise time

The relationship is quite linear; the longer rise times lead to later P-centres. The rise time of the stimulus is affecting the signal's P-centre.
The slope of the regression lines of the plots on Figures 9.8 9.9 and 9.10 give an indication of the relationship between the rise time of a signal and the P-centre. These slopes can be compared to establish the variation in the effect of rise time on P-centre as a factor of different speech sounds.

Table 9.10 below shows the regression equations of the rise time/P-centre plot for each set of speech stimuli; the gradients (slope values) are emboldened.

From Table 9.10, it can be clearly seen that the slope of the ramped "sha" stimuli is different from the slopes of the "wa" and "ae" stimuli. The gradients of the "wa" and "ae" stimuli (-0.352 and -0.346 respectively) are very similar to one another. This relationship indicates that for these stimuli, for a 10ms increase in rise time, there is an approximate 3.5ms shift of the P-centre location away from the onset of the signal.

speech stimuli	regression equation	r^2 value
"sha"	PC = -149 - **0.0339** rise time	52.9%
"wa"	PC = -2.80 - **0.352** rise time	61.6%
"ae"	PC = 0.90 - **0.346** rise time	91.2%

Table 9.10 linear regression equations of P-centre/measured rise time plots for each set of ramped speech stimuli (the equation fitted in linear regression is $y = c + \alpha x$)

For the "sha" stimuli, the gradient is shallower (-0.0339); therefore for a 10ms increase in rise time of the signal, there is an approximate 0.34ms shift in P-centre away from the onset of the signal.

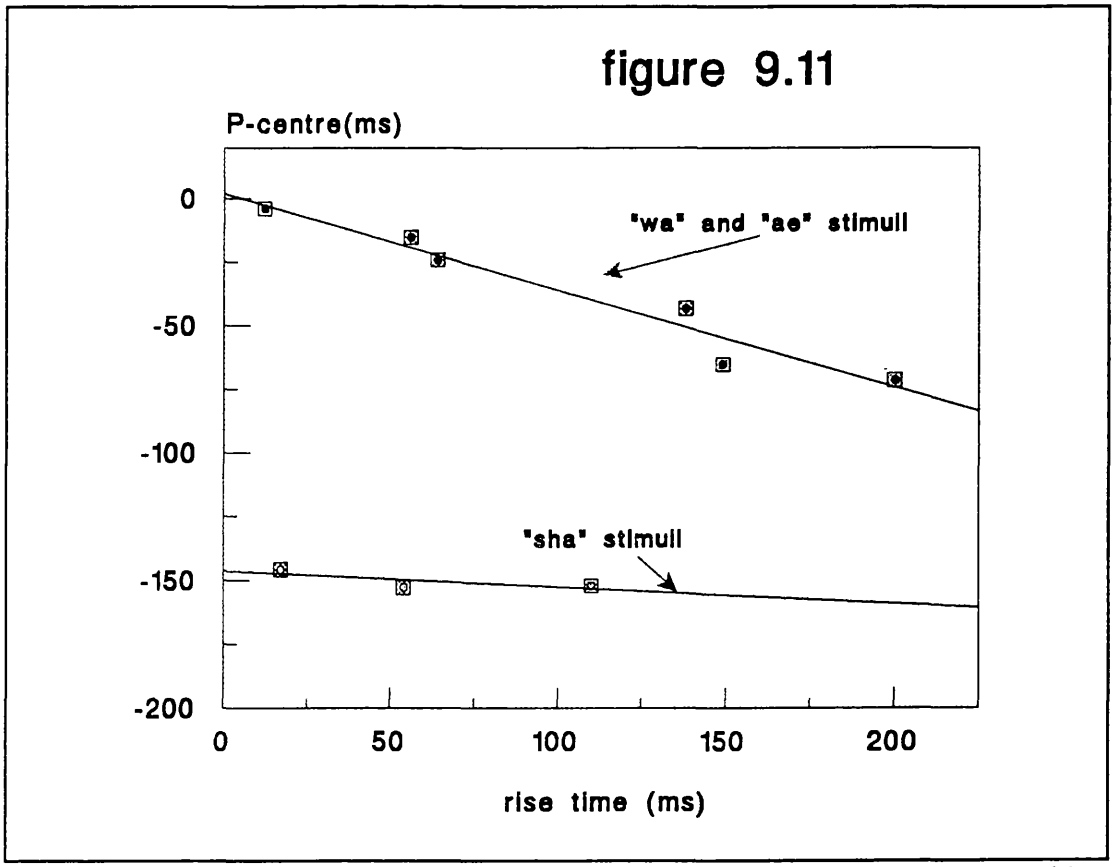


Figure 9.11 P-centres of "wa" and "ae" stimuli plotted against rise time with regression line fitted; P-centres of "sha" stimuli plotted with regression line

Figure 9.11 shows the P-centres for the "wa" and "ae" stimuli together, against the measured rise time of the signals. The regression line is plotted (the model fitted is $y = c + \alpha x$).

The regression of the P-centres of the "wa" and "ae" stimuli against the predictor measured rise times had the equation:

P-centre = 2.08 - 0.380 measured rise time

The predictor **measured rise time** was significant ($t_{1,4} = -8.75$, $p < 0.05$).

There is thus a very significant effect of the rise time of a signal upon the P-centre of the signal, and this effect is of the same magnitude for the "wa" and the "ae" speech stimuli.

Also plotted on Figure 9.11 is the P-centre/rise time plot of the "sha" stimuli - this shows a much shallower relationship that is clearly different from that of the "wa" and "ae" stimuli.

9.13 Implications for models of P-centre location

There is a dissociation in the effects of ramping the rise times of speech stimuli on the P-centres of the stimuli; this dissociation is due to the contents of the speech signal. Ramping a portion of frication has little effect on P-centre location; ramping a semi vowel has an effect on P-centre location; ramping a vowel has a very similar effect.

The results of the "sha" ramping would be generally predicted by Marcus's model; this manipulation does not affect the vowel onset and therefore not the P-centre. Marcus's model might predict the effect of the ramping upon the "wa"

stimuli, since this could affect the vowel onset as defined in his model, and thus shift the P-centre. Marcus's model would not expressly predict the effect of ramping upon the vowel, since by definition the signal starts with a vowel. His criteria for vowel onset (peak increment in mid band spectral energy) might vary with the ramping, and thus the P-centre might vary. If this prediction is fulfilled, then the status of Marcus's vowel onset criteria should be threatened. It might be a vital determinant of P-centre location - but does it represent vowel onset, or is it a quantification of amplitude changes within a specific frequency bandwidth?

9.14 Marcus model predictions

Using the Marcus model implementation described in Experiment 3, predictions were calculated for the P-centres of all the sets of ramped stimuli. These are shown in the Table 9.11.

speech stimuli	measured rise time (ms)	Marcus model prediction (ms)
"sha"	17.0	-328.0
	54.0	-328.0
	110.0	-328.0
"wa"	138.0	-258.3
	149.0	-258.3
	200.0	-258.3
"ae"	12.0	-53.3
	56.0	-59.3
	64.0	-77.3

Table 9.11 Marcus model predictions of P-centre locations for speech stimuli ramped to different rise times

Table 9.11 shows that the Marcus model predictions are constant for the "sha" and "wa" stimuli. The ramping of these stimuli does not affect the vowel onset as defined by his model. The Marcus model predictions for the "ae" stimuli do shift away from the onset with the rise time of the signals; this indicates that the 'vowel onset' can capture aspects of the amplitude distribution of a signal.
The Marcus model predictions were used in a linear regression fitting the equation:

$$y = c + \alpha x$$

where y=P-centre and x=Marcus model predictions

This was to test the extent to which the Marcus model predictions accounted for the observed variation in P-centres due to ramping the onset. The calculated P-centres were regressed against the predictor **Marcus model values** (shown in Table 9.11).

The regression had the equation:
P-centres = 23.7 + 0.456(Marcus Model Predictions)

The predictor **Marcus model values** is significant ($t_{1,7} = 5.18$, $p < 0.05$).

The Marcus model therefore accounts for a significant amount of the observed variance (76.4%). This is despite the model making no discrimination between the P-centres of the "sha" and the "wa" stimuli.

The plot on Figure 9.12 of measured P-centres of all the stimuli against the Marcus model predictions reveals the source of this discrepancy. The Marcus model accounts for the overall differences between the P-centres of the three types of speech - which vary a great deal. Thus the P-centres of the "ae" stimuli vary between -3.75 and 24.0 ms; those for the "wa" stimuli between -43.25 and -71.5 ms; and those for the "sha" stimuli between -148.5 and 152.0

ms. It is this gross range of P-centres that the Marcus model predicts. The smaller, within speech type differences, which are due to ramping the stimuli, are not accounted for by this model.

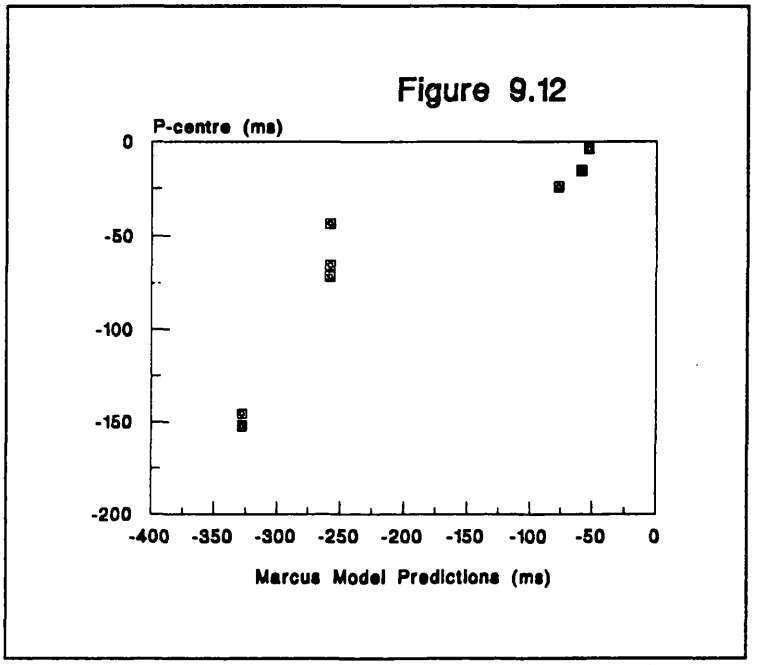


Figure 9.12 P-centres of all ramped stimuli plotted against Marcus Model P-centre predictions

9.15 Implications for Marcus's Model of P-centre location

Marcus's model does not capture all the aspects of a signal which determine its P-centre. A parameter of intensity change needs to be incorporated. Marcus's model implicitly contains a parameter of intensity change with a mid range spectral bandwidth - the vowel onset is defined as the peak increment in energy between 500-1500Hz. Marcus's model is structured around the division of the signal into two around vowel onset. Marcus's model thus takes as its most important feature an attribute of the rise time of the amplitude within this bandwidth.

Could this issue be considered from a different aspect, one which emphasized the role of frequency specific amplitude changes? Instead of Marcus's model representing P-centres as a combination of the durations of the two portions of a signal, and thus being a global account, his model could be regarded as stating that P-centres relate to attributes of the acoustic vowel onset, and thus being a local account. From this perspective, Marcus's model predicts that a longer prevocalic consonant leads to a later P-centre because the longer the prevocalic consonant, the later the vowel onset is.

Following this thread, it has been well determined that P-centres are not marked by the vowel onset *per se* (Morton, Marcus and Frankish 1976, Marcus 1981). Others have found (Cooper Whalen and Fowler 1986) that a 1ms alteration of the prevocalic consonant duration of a speech item results in a 1ms shift in the measured deviation from isochrony (implying a shift in P-centre of this amount). Such a 1:1 relationship implies that the prevocalic consonant affects P-centre only in as much as it delays the vowel onset.

If vowel onset alone does not predict P-centres, then an extra level of information could be provided by considering the amplitude profile at the onset of the vowel. The results of the current experiment suggest that the parameter of vowel onset rise time is indeed affecting P-centre location. Perhaps Marcus's model could be inverted, and added to, and a model of P-centre location which is *local* and *concerned only with the rise time of a subset of the spectral contents of a signal* could be developed.

9.16 Frication ramping and P-centres - discrepancy with previous findings?

Howell in 1984 found that there was an effect of ramping the fricative "cha" on the settings made by subjects in a type of rhythm setting task. This experiment showed a much reduced effect of the same manipulation upon P-centre location. Are these results incompatible?

There are differences between the procedures used that reduce the possibility of comparison. Howell did not use a full rhythm setting task, and therefore P-centre calculations could not be made from his data. The calculation of P-centres rests upon the comparison of each stimulus with several others, several times; large differences between two particular settings tend not to be reflected in large P-centre differences, since the fit is a conservative estimate, and one based upon *mean* settings. Thus there is a difference in the P-centres of the ramped "sha" stimuli, but it is a small one, especially when compared to the difference caused by the same manipulation on other stimuli. These experimental results are different to Howell's in that they go further, and show larger differences between types of stimuli than within some stimuli.

9.17 Conclusions

Ramping the onset of speech sounds, and thus changing their measured rise times, alters their P-centres. The shift caused by this manipulation is away from the onset of the signal. Longer rise times lead to later P-centres.
This shift in P-centres caused by the differential ramping was not equal for all speech sounds. Ramping the onset of a semi vowel ("wa") and a vowel ("ae") had effects that were linear and very similar for both speech types. Ramping the onset of a fricative ("sh") without affecting the prevocalic consonant duration

had a very slight effect on P-centre location. This implies that ramping the onset of a signal is dependent upon the frequency content of the onset portion of that signal. Ramping high frequency noise has a very reduced effect on the P-centre location. Ramping lower frequency, more sonorous speech sounds such as vowels and semi-vowels has a large effect on P-centre location. Marcus's model predicts the gross difference in P-centres between speech types, but not the effects of ramping the onsets.

Chapter Ten
Experiment 7
A Comparison of the effect of ramping the stimulus onset or offset on P-centre location - are onset events more important?

Abstract
The previous experiments have indicated that the onset characteristics of a signal affect the P-centres of the signal. In contrast, the amplitude characteristics of the offset of a signal do not affect the P-centre. This assertion was tested by performing the same manipulation on the *onset* and the *offset* of a signal, and comparing the resultant P-centres. A synthetic vowel with a constant amplitude and spectral profile was used; in one set of stimuli the onsets were ramped, in the second set the offsets were ramped to the same degree. the results indicated that ramping the onset of a vowel affects the P-centre, while ramping the offset has a negligible effect.

10.1 Introduction

The image of P-centres that the experiments so far have outlined is that they are principally influenced by the onset amplitude characteristics of a subset of frequencies, which can be considered as corresponding to qualities of the vowel onset. This is based on several experimental findings:

1) amplitude change at the onset has more effect on P-centre location than concurrent changes at the offset, when the speech signal is infinitely peak clipped (Experiment 4).

2) ramping the onset of speech sounds alters the P-centres (Experiment 6).

3) ramping the onset of some speech sounds (eg. vowels) has more effect than ramping others such as fricatives (Experiment 6). This implies that there are frequency dependent attributes of rise time on P-centre location.

4) amplitude change at the offset alone does not alter P-centre location (Experiment 5); increasing the amplitude of a /t/ burst does not shift the P-centre.

The hypothesis that offset events are not as relevant as onset events as defined by rise times has been supported by the evidence so far. However, Experiment 5, where offset events alone were manipulated, contained only manipulations to a /t/ burst. This segment contains high frequency noise - precisely the spectral contents, the amplitude manipulation of which has little effect on P-centre (as indicated by the results of Experiment 6). What would be the effects of manipulating the amplitude of a speech segment whose spectral contents have been implicated in the P-centre location of the signal?

10.2 A pilot study

The clear experimental manipulation was to test the effect of increasing the amplitude of a vowel sound at the offset of the signal, compared to changes at the onset. A pilot study was performed where the amplitude was doubled for a 50ms portion of synthesized vowel sound at the onset, middle and offset. the duration of the sound was kept constant at 200ms. A synthesized vowel sound was used so that the spectral contents could be controlled over time. This meant that altering the offset amplitude had the same spectral consequences as altering the amplitude at the middle or onset of the signal. The rise time of the signal was constant at 10ms. Two subjects set these stimuli in a dynamic

rhythm settings task, along with the reference sound. The A - A interval was 1250ms.

An unforseen consequence of this manipulation was that the increase in amplitude, when it occurred at the middle or end of the vowel, resulted in a 'click' in the signal. This is because the rate of change of the increase in amplitude was the same as that for the onset of the signal. The click gave the signal a double beat characteristic, implying that a signal containing two rapid increases in amplitude has the attributes of two signals or acoustic events. This 'doubleness' was reflected in the subjects settings. They made widely varying settings which on analysis appeared to vary around two intervals per stimulus. There was evidence that both subjects were setting the stimuli with the amplitude increases at the middle and offset at two different intervals per stimulus. The double click of the signals was presenting the subjects with two rhythmic centres, or P-centres, to base their rhythmic judgements upon. Thus the subjects made their settings based upon the perceptual events associated with the *onset* of the signal, or on those associated with the *amplitude increment* in the signal. This was not the case for the signal with the increase in amplitude at the onset; this only contained one rapid increase in amplitude, resulting in one rhythmic centre to the signal. The principle described here is not however pure speculation; stimuli were presented in a talk by A. Bregman[1] which had similar attributes.

A continuous sine tone signal was played; it had increases in amplitude which were very gradual (had long rise times and thus a slow rate of change of amplitude with time). This led to the perception of one signal which increased and decreased in intensity. If the rise times of the energy increments were drastically shortened (so the rate of change of amplitude with time was fast) the

[1] Stimuli presented in a talk at the Institute of Acoustics Speech Group Meeting, "Perceptual separation of speech and other complex sounds", Feb. 1991, University of Sussex

single amplitude varying signal 'disappeared' and instead the perception of two signal arose; one continuous and unvarying in amplitude; the second consisting of an intermittent 'click' of the same pitch as the continuous tone. Thus the rate of change of amplitude in a continuous tone affected whether one or two signals were heard. This phenomena, in the context of a rhythm setting task, resulted in no single P-centre for the experimental stimuli.

The basic assumption underlying all dynamic rhythm setting tasks - that the subject uses the P-centre of a signal to set that signal to a rhythm - cannot be made for this study. The double beat of two of the three experimental stimuli, as reflected in the subjects settings, meant that these settings could not be used to determine P-centres for the stimuli.

There was problem with correcting for this double beat phenomena. A rise time for the increase in energy that did not lead to the perception of a second beat would have to be found, and this same rise time used for the onsets of the signals. This was not a realistic option, since it would require a great deal of experimentation simply to choose the exact manipulations which could then be used experimentally.

10.3 Are onset events more important - a test

The issue remained of how to test the hypothesis that onset events are more important than offset events when tested using a vowel. The previous experiment showed that the P-centre of a signal varied with the stimulus rise time, for vowel and semi-vowel stimuli. This was shown with naturally produced speech sounds. In this experiment the same manipulation was performed upon the onset of a synthesized vowel - and also upon its offset. The effects, not of increasing, but of reducing the energy at the signal onset and offset in terms of the rise/decay time, were compared.

Increasing the rise time of a vowel should shift the P-centre away from the onset as was demonstrated in Experiment 6. If the offset events are equally important in determining the P-centre, then increasing the decay time of the offset should shift the P-centre towards to onset of the signal. If the offset events are less important in determining the P-centre, then the P-centre should not vary with the decay time to the same degree, if at all.

The effects of ramping the onset or offset of a vowel on the P-centre location can be compared by analysis of the resultant amount of ramping/P-centre location relationship. Thus the effect of ramping the stimulus onset and offset on P-centre was explicitly tested in an experiment. A synthetic vowel with a constant duration and decay time was ramped at the onset to four different rise times. An identical synthetic vowel with a constant rise time and duration was ramped at the offset four different decay times. The ramping levels were the same in the two conditions. The vowel was constant in pitch to ensure no variation in the effects of ramping. P-centres were determined for the resulting eight stimuli in two dynamic rhythm setting tasks. The effects of ramping the onset and offset of signals on the P-centres of the signals was explicitly compared.

Hypotheses, based on the previous experimental results can be stated:

1) That ramping the onset of the synthetic vowel will alter the P-centre of the signal. Longer rise times will lead to later P-centres.

2) That ramping the offset of a synthetic vowel will not alter the P-centre of the signal as much as ramping the onset, if at all.

10.4 Method

10.4.1 Stimuli

The stimuli used were taken from those described by Kuhl and Miller (1978). These were synthesized at the Haskins laboratories on the parallel-resonance synthesizer in accordance with the parameter files developed by Abramson and Lisker (1970). The stimuli used in the current experiment were edited from the steady state portion of the "ah" vowel, such that the onset rise time was 0ms, decay time was 0ms, and the fundamental frequency was a constant 114Hz. A spectrogram of this vowel is shown in Figure 10.1.

A synthesized vowel was used to ensure that:

i) The onsets and offsets could be manipulated with a linear ramp such that the applied ramp was similar to the resultant rise time. A naturally produced vowel, with a ramped onset/offset would result in levels of rise time/decay time that were not regularly varied.

ii) The amplitude was constant over the duration; there were no amplitude variations that would lead to irregular rise/decay time variation over the stimuli.

iii) The spectral contents were constant over the duration of the vowel - thus rise/decay time manipulation would not be differentially affecting frequency information.

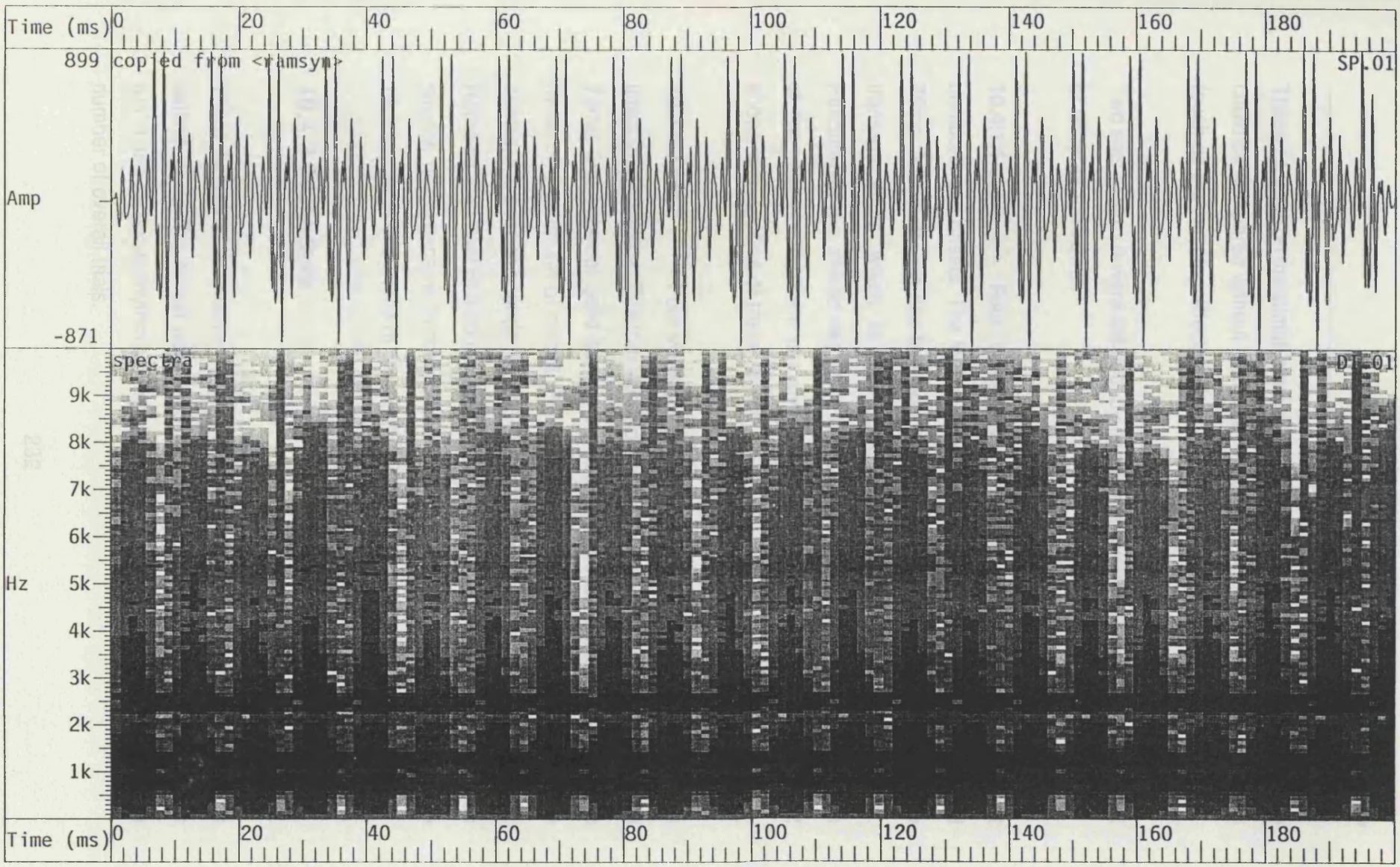


Figure 10.1 Oscillogram and Spectrogram of synthetic vowel "ah"

These controls on the stimuli manipulations meant that the onset/offset ramping could be compared without spurious effects distorting the experimental results (such as the ramping affecting the onset more than the offset).

Two sets of stimuli were created, by applying a linear scaling ramp to the onset or offset of the vowel.

10.4.1.1 set 1 Four vowels, with different rise times and constant durations and offsets. The four levels of rise time were 5ms, 25ms, 50ms and 75ms. A level of 0ms rise time was not used as such an abrupt onset lead to transient clicks which in turn could have caused experimental errors. Perceptually the altered rise times led to degrees of gentleness at the onsets of the vowels; they were perceived as reasonably speech-like. Figure 10.2 shows oscillograms of these stimuli.

10.4.1.2 set 2 Four vowels, with different offset times and constant rise times and durations. The four levels of decay time were 5ms, 25ms, 50ms and 75ms. 0ms was not used so that the results could be explicitly compared to those of the first set of stimuli. Perceptually this manipulation had noticeable effects; the speech sounds took on an 'instrument' quality, rather like a piano. Ramping the offset of a sound does affect its perceptual qualities. The speech sounds never became entirely unspeech-like, but did sound unusual. Figure 10.3 shows oscillograms of these stimuli.

10.4.2 Procedure

These two sets of stimuli - 16 tokens in all - were used in dynamic rhythm setting tasks. The stimuli were used within each set - that is set 1 stimuli were run in one complete rhythm setting task, and set 2 in another. This reduced the number of overall trials.

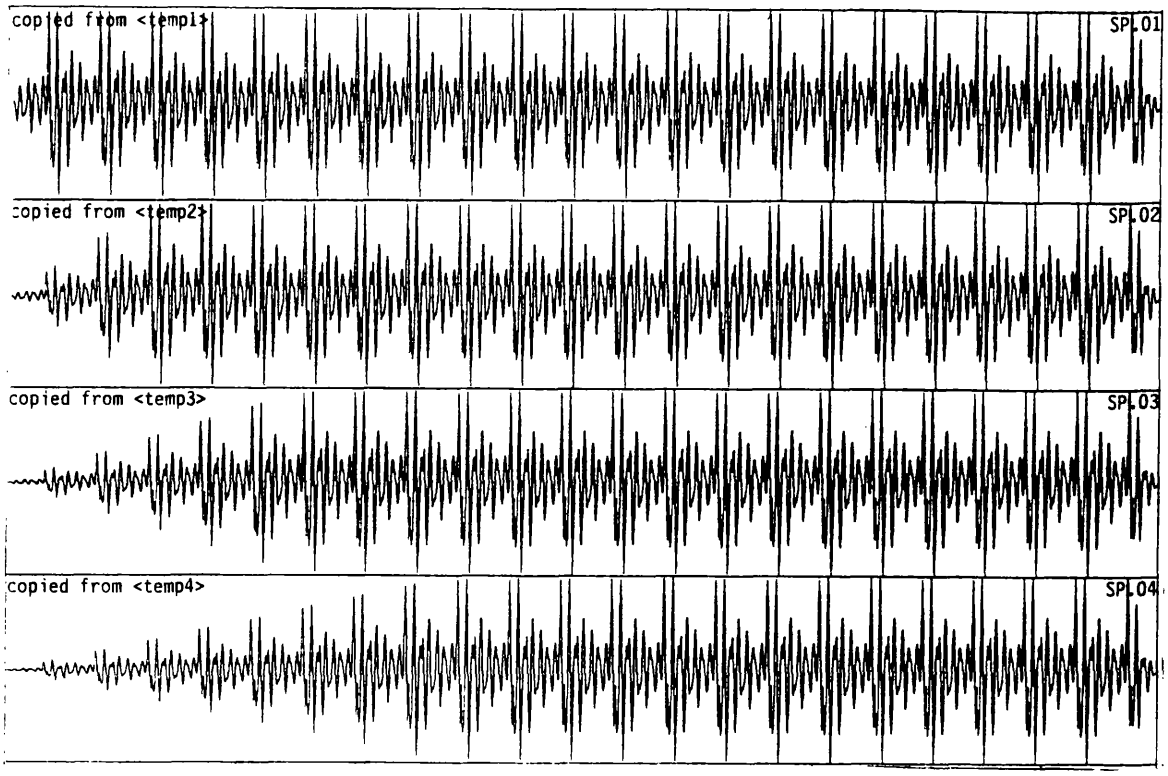


Figure 10.2 Oscillograms of stimuli ramped at onset to 5ms, 25ms, 50ms and 75ms

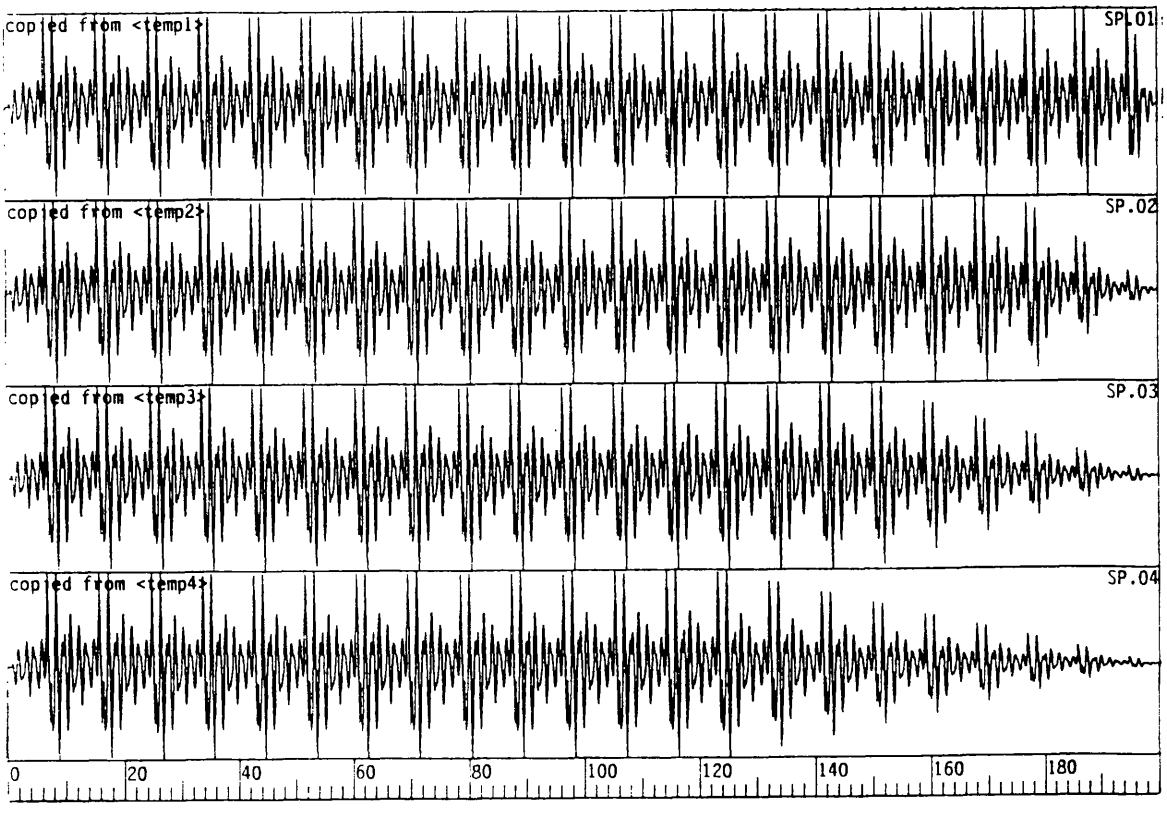


Figure 10.3 Oscillograms of stimuli ramped at offset to 5ms, 25ms 50ms and 75ms.

Included in each task was the reference sound. This meant that the results could be compared. In addition, stimulus 1 in each experiment was identical (5ms onset, 5ms offset, 200ms duration). This increased the strength of the comparison. The P-centres for the two sets of stimuli were based on dynamic rhythm setting tasks in which two of the stimuli were identical. The A - A interval was 1250ms.

10.4.3 Subjects

Two experienced subjects performed the experiment - SHS and SKS. Subject SKS was not naive to the experimental manipulations.

There were two sets of four stimuli, plus the reference sound. The experiment consisted therefore of two sets of 25 stimuli combination blocks, with 5 trials in each block. The trials that were identical to both sets of stimuli were only performed once each, to reduce the overall number of trials. Thus the combination blocks of stimulus 1 - stimulus 1, reference sound - reference sound, and both stimuli 1 - reference sound combinations were performed once only. This reduced the overall number from 250 individual trials to 230 trials overall, per subject.

10.5 Results

The means and standard deviations of the settings the subjects made for the stimuli combinations, for both sets of stimuli are shown in the Tables 10.1 and 10.2 below.

Table 10.1 shows the mean intervals set for each stimulus combination for the stimuli with the ramped onsets (RT). Since the A - A interval is 1250ms, any

deviations from 625ms indicates a difference in the P-centres of the stimuli concerned.

stimuli	5ms RT	25ms RT	50ms RT	75ms RT	ref
5ms RT	624.30 7.21	620.00 6.96	612.12 11.32	607.80 16.04	625.78 8.78
25ms RT	634.00 12.48	622.56 10.62	618.67 10.19	608.90 18.05	633.44 14.91
50ms RT	631.57 14.62	629.37 12.52	623.87 8.63	615.20 11.42	644.89 14.91
75ms RT	644.11 13.76	634.67 12.29	632.50 7.89	623.80 9.03	635.50 11.04
ref	621.44 12.70	629.87 15.68	613.63 13.96	613.89 11.14	628.00 9.48

Table 10.1 of mean and standard deviations of final interval settings for stimuli with different rise times - RT (set 1)

Table 10.2 shows the mean intervals set for each stimuli combination for the stimuli with the ramped offsets. Again, deviations from 625ms indicate that the stimuli have different P-centres.

On both sets of data, the trials where the same sound is set against itself (shown in the shaded cells) have mean intervals which are around 625ms. This indicates that the subjects were performing the task correctly - that is, they could set one repeating sound to an isochronous rhythm. In addition the standard deviations are small. This shows that the subjects were reasonably consistent in their settings, and not merely performing randomly. Comparing the two tables further, differences appear. For Table 10.1 intervals between stimuli with a short rise time (RT) and a long rise time tend to vary from 625ms. This forms a general pattern over all the intervals. This indicates a difference in the P-centres due to the alteration of the rise times. The same does not seem to

be the case for the intervals shown in Table 10.2, most of which are centred around 625ms.

stimuli	5ms DT	25ms DT	50ms DT	75ms DT	ref
5ms DT	same as Table 10.1	624.00 9.66	622.50 10.13	631.11 9.31	same as Table 10.1
25ms DT	627.50 7.69	617.9 9.80	625.11 9.49	627.10 10.82	622.22 15.00
50ms DT	620.17 12.48	626.10 7.05	624.67 8.03	624.13 6.08	627.50 7.55
75ms DT	612.89 14.71	627.67 5.17	619.20 8.26	624.13 6.08	627.50 7.55
ref	same as Table 10.1	636.60 15.62	624.80 16.02	623.50 25.00	same as Table 10.1

Table 10.2. Mean and standard deviations of the final interval settings for all stimuli combinations for stimuli with different decay times - DT (set 2)

The significance of these stimuli combinations on these intervals (and any subject differences) was established by regressing the absolute deviations from isochrony against the stimuli combinations and the subjects. This gives a measure of the absolute amount by which a subject must shift a signal from absolute isochrony to acheive perceptual regularity.

Tables 10.3 and 10.4 below show the means and standard deviations of the absolute deviations from isochrony of the subjects' settings for the two sets of stimuli. This gives a measure of the amount of shift from perfect isochrony the subjects made to acheive perceptual isochrony.

As noted in Tables 10.1 and 10.2. the trials where the same sound is set against itself (shaded cells) have smaller mean deviations from isochrony - the subjects set these stimuli to an isochronous rhythm. The other cell contents Table 10.3 (ramped onset stimuli) show larger mean deviations from isochrony,

indicating that these stimuli have differing P-centres. The cell contents of Table 10.4 (ramped offset stimuli) do not vary as markedly from the same stimuli means - indicating that these stimuli do not have P-centres that are very different from one another.

stimuli	5ms RT	25ms RT	50ms RT	75ms RT	ref
5ms RT	5.700 4.27	6.556 5.318	14.125 9.463	18.400 14.485	7.667 3.428
25ms RT	11.00 10.73	8.000 6.892	8.111 8.652	17.500 16.534	10.444 9.153
50ms RT	14.125 9.463	10.125 7.846	7.375 3.701	12.000 8.781	20.778 13.479
75ms RT	19.111 13.761	17.500 16.534	12.000 8.781	6.400 6.132	11.900 9.327
ref	10.667 6.892	11.875 10.575	12.875 12.380	13.556 7.502	8.400 4.624

Table 10.3 means and standard deviations of absolute deviations from isochrony for all stimuli combinations for ramped onsets (ms)

stimuli	5ms DT	25ms DT	50ms DT	75ms DT	ref
5ms DT	as Table 10.3	7.667 5.315	8.500 5.210	7.889 7.656	as Table 10.3
25ms DT	6.500 4.378	9.900 6.57	7.889 4.485	8.500 6.451	12.556 7.502
50ms DT	11.500 5.01	5.500 4.17	6.333 4.416	8.000 6.595	5.600 4.402
75ms DT	16.333 9.014	4.667 3.202	7.800 6.161	4.875 3.271	6.700 3.743
ref	as Table 10.3	16.000 10.446	12.600 8.959	20.500 12.669	as Table 10.3

Table 10.4 means and standard deviations of absolute deviations from isochrony of settings subjects made with ramped offset stimuli

The significance of effect of the stimuli on the subjects settings was tested statistically. The standard deviations were not too large, so the data was not transformed. Both of these sets of deviations were regressed against the predictors **stimuli combinations** and **subjects** using multiple regression, fitting the equation:

$$y = c + \alpha(x_1)$$

where y=deviations from isochrony, x1=stimuli combinations

This was to determine for each whether the combinations were significantly affecting the settings, or whether the subjects were a significant source of variation.

10.5.1 Ramped onset stimuli (set 1)

The regression of the absolute deviations from isochrony for the onset ramped stimuli against the predictors **stimuli combinations** and **stimuli** had the equation:

absolute deviation = 9.62 + 0.338(stimuli combination)

The predictor **stimuli combination** was significant ($t_{2, 222} = 4.08$, $p < 0.05$).

To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was significant ($t_{2,222} = -4.47$, $p < 0.05$). The two subjects therefor differed significantly in the settings they made. The predictor **stimuli combinations** accounted for 6.2% of the variance; the predictor **subject** accounted for 8.2% of the observed variance.

10.5.2 Ramped offset stimuli (set 2).

The absolute deviations set for the ramped offset stimuli were regressed against the predictors **stimuli combinations** . The equation was:

absolute deviation = 4.19 + 0.191(stimuli combination)

The predictor **stimuli combinations** was significant ($t_{2, 229} = 2.99$, $p < 0.05$). To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was not significant ($p>0.05$). The two subjects were thus consistent with each other in their settings for the offset ramped stimuli.
Thus for the stimuli ramped at the offset, there is a significant effect on the absolute deviations from isochrony of the stimuli combinations. Unlike the previous regression, however, the subjects are not a significant source of variance.

10.6 Discussion of regression results

To calculate P-centres from the patterns of intervals set by the subjects in dynamic rhythm setting tasks, it must be assumed that the subjects are consistent with one another. If this assumption cannot be made, then the P-centres cannot be calculated until the subject differences are accounted for.

The same subjects took part in both parts of this experiment. Differences between then due to hearing problems or inability to perform the task would thus be expected to be apparent on both regressions; instead there is a difference on just one of the sets of stimuli - those with the ramped onsets.

There was no clear pattern of differences when the subjects mean deviations from isochrony were plotted against stimuli combination. To establish the nature of the difference therefore, the standard deviations of each subjects' individual

mean absolute deviations from isochrony are plotted against the mean deviations. This will show is the distribution is skewed (if SD increases with mean) and if there is a difference in the ranges of the means and SD's for each subject. These SD/mean plots are shown on Figure 10.4.

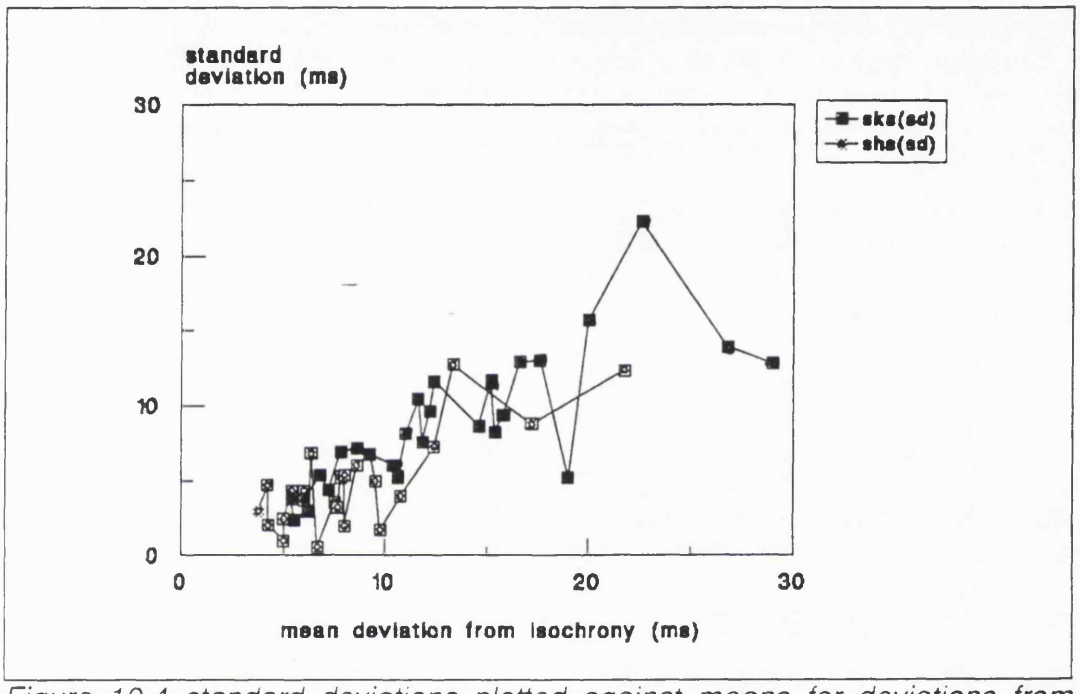


Figure 10.4 standard deviations plotted against means for deviations from isochrony set by both subjects SHS and SKS

Both plots show increased standard deviations as the means increase, which is indicative of a skewed distribution (and hence the transformation described earlier). In terms of the subject differences, it can be seen that subject SHS set smaller means, with smaller standard deviations, than subject SKS. The SHS plot shows lower y axis values, indicating smaller standard deviation values; the plot for SKS extends further along the x and y axis indicating that the range of SKS's responses was wider and with greater variation than those of SHS. It would seem therefore that subject SKS performed the dynamic rhythm setting task less accurately than subject SHS. This suggests that the subject

differences are a result of subject SKS showing considerably more variance in her settings than subject SHS, as reflected in the standard deviations of their settings. This means that the subject SHS was 'better' at this task, in terms of being internally consistent. This was confirmed by a subjective account from SHS (the naive subject), who stated that he preferred to not complete a trial, rather than make a setting he was unhappy with.

There was no significant effect of the subjects on the settings made with the ramped offset stimuli. This implies that, for whatever reason, the non-naive subject made more variable responses when setting the ramped onset stimuli to an even rhythm. This could be attributed to fatigue or lack of attention - certainly the motivation of this subject was not in doubt. There was no evidence, here or elsewhere, that this subject either could not do the task, or was at some level adjusting her settings to ensure a positive experimental result.

Since there was no evidence for a consistent differential effect of the manipulation upon the two subjects, their results were pooled for both sets of stimuli, and P-centres calculated, with one reservation. The P-centre algorithm, as described in Chapter 4, compensates for 'over turning' of the knob when making the settings, and thus should reduce some of the error due to subject variation. In addition, a further check was made, by comparing the measured relationship between onset ramping and P-centre, and that observed in Experiment 6 between ramping the onset of a natural vowel and the P-centre of that vowel. If the P-centres of the ramped synthetic vowel are reasonable, the two onset ramp/P-centre slopes should be of similar orders. If there is a different in the slope of the relationships, this would indicate that subject differences had affected these results, and that the experiment should be repeated with different subjects.

10.7 P-centre calculations

The mean pooled intervals set in the two stimuli conditions were used to calculate the P-centres of the stimuli; these were adjusted so that the reference sound P-centres were equal to zero. The P-centres are shown in the Table 10.5 below. The P-centres were analyzed with respect to the amount by which the synthetic vowel was ramped at the onset and offset, since there was no natural ramping of the signal and it originally had a completely rectangular amplitude envelope.

Amount of ramping (ms)	P-centres (ms)	
	onset ramped	offset ramped
5	+0.4	-1.0
25	-4.2	+2.4
50	-10.2	+1.0
75	-16.0	+3.6

Table 10.5 Calculated P-centres for vowels with different rise/decay times.

The "+" signs indicate that the P-centres of these stimuli are earlier than the P-centre of the reference sound, since all the P-centre values are adjusted such that the reference sound P-centre equals zero.

From Table 10.5 it can be seen that ramping the onset of a signal is having a clear effect on the P-centre of that signal. The greater amount of ramping results in later (more negative) P-centres. Ramping the offset has a less clear effect, though there is a rough shift of the P-centres towards the onset of the sound with increased ramping of the offset.

This was tested statistically by regressing the P-centres of the signals against the amount of ramping for both sets of stimuli. This was done using linear regression, fitting the equation:

$$y = c + \alpha x$$

where y=P-centre and x=amount of ramping

10.7.1 Onset ramped stimuli

The P-centres for the onset ramped stimuli were regressed against the predictor **amount of ramping**. The regression had the equation:

The regression equation for the onset ramped stimuli is:
P-centre = 1.60 - 0.235 onset ramp

The predictor **amount of ramping** was significant ($t_{1,2}$= 782,18, $p < 0.05$).

There is therefore a strong linear relationship between the amount by which the onset is ramped and the P-centre of the vowel. The slope of the relationship is (-0.235). Thus the longer the onset ramp, the later the P-centre.

10.7.2 Offset ramped stimuli

The P-centres for the offset ramped stimuli were regressed against the predictor **amount of ramping**.

The regression equation for the offset ramped stimuli is:
P-centre = -0.36 + 0.0500 offset ramping

The predictor **amount of ramping** was not significant ($p > 0.05$).

Thus there is a non significant relationship of the offset ramping upon the P-centre of the stimuli. The slope of the relationship is (+0.05); slightly positive.

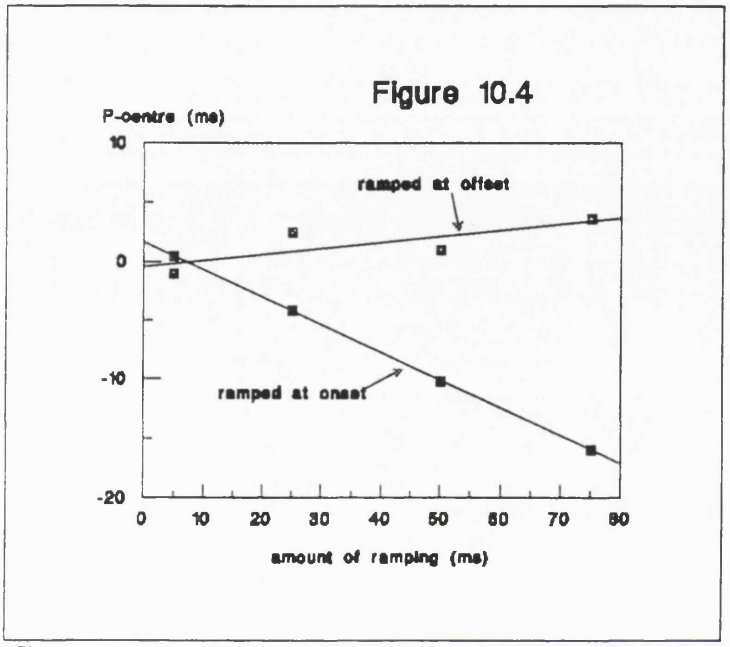


Figure 10.5 P-centres of onset ramped and offset ramped sets of stimuli, plotted against overall amount of ramping

Figure 10.5 shows the P-centres of the two sets of stimuli, plotted against the total amount of ramping. The relationships outlined by the regression results are reinforced here; the P-centres of the onset ramped stimuli become more negative as the amount of ramping increases. The P-centres of the offset ramped stimuli become slightly more positive as the amount of ramping increases, but this relationship is not significant.

1) Systematically increasing the rise time of a signal with a constant offset and duration shifts the P-centre in a linear manner. The gradient of the relationship is (-0.235). The relationship is highly significant, with the amount of ramping accounting for all the observed P-centre variation.

2) Systematically increasing the decay time of a signal with a constant onset and duration does not shift the P-centre in a linear manner. The observed

variance in P-centres, which is low, is not significantly accounted for by the variation in decay time.

The slope of the onset ramped stimuli with ramping was (-0.235); the slope of the onset ramped natural vowel with measured rise time in Experiment 6 was (-0.346). The two slopes are of similar orders, both are negative and less than one. The slope calculated in this experiment is somewhat shallower (that is, closer to zero) than that from Experiment 6; this indicates a slightly less strong relationship between ramping and P-centre for these onset ramped stimuli. This means that the subjects' results do not need to be discarded for being too unusual or extreme; the overall result is more conservative than that of Experiment 6. There is no method of objectively measuring P-centre locations that does not rely on subjects' settings, therefore this result will be accepted.

10.8 Discussion.

The results of part of Experiment 6 showed that the rise time of a speech sound affected the P-centre of the signal, unless the onset of the signal was a low sonority phoneme such as a fricative. The maximum effect of ramping the onset of a signal was when the speech sound onset was a vowel or a semi-vowel.

This finding was replicated in the current experiment. Altering the rise time of a synthetic vowel with a linear ramp systematically shifts the P-centres of the signals. The slope of the regression of P-centres against rise times is (-0.235) for the synthetic vowel /a/, and (-0.346) for the natural vowel /ae/. The slopes are not identical; but they are similar enough to conclude that the parameter of rise time manipulation has a similar effect on both types of speech stimuli, despite one being synthetic, and their being used in different experiments with different subjects. This effect on synthesized speech, which sounded clearly

artificial, is highly suggestive that this phenomena is not speech specific but is instead applicable to all acoustic signals (cf. Gordon's 1987 model of musical note perceived attack time in Chapter Two).

Ramping the decay time of the same synthetic vowel does not systematically shift the P-centres of the signals. This indicates that, although perceptually salient, the decay time of a signal does not affect its P-centre. This is therefore congruent with the results of Experiment Five in which it was determined that an increase in /t/ burst at the offset of a signal does not affect the P-centre in production or perception.

These results - showing an effect of altering the onset amplitude characteristics of a signal on its P-centre location, but little effect of ramping the offset of the signal on its P-centre - afford the controversial interpretation that this is because the rise time of a vowel determines its P-centre. Thus the offset manipulation would have no effect on P-centre location since it does not alter the hypothesized P-centre determinant, vowel onset rise time. Likewise, increasing the amplitude of a syllable final /t/ burst would not alter the P-centre of a syllable for the same reason (see Experiment Five). This statement is controversial because it negates one of the main P-centre parameters - the duration of a signal - and thus is not a global account of P-centre location, rather a local one.

How can this approach be justified? As outlined earlier, it is not wholly different from Marcus's model in that it stresses the onset of the vowel. It goes further, in that it addresses acoustic attributes of the vowel onset, specifically rise time. Also it does not necessarily explicitly define vowel onset, but could solely be concerned with amplitude changes within certain frequency bandwidths that might well correspond with vowel onset, but also other sonorous speech sounds ("w", "y" etc).

Previous work on P-centres (for instance Marcus 1981, Cooper, Whalen and Fowler 1986, 1988, Fox and Lehiste 1987) has focused on the duration of the prevocalic and post vowel onset portions of a syllable and their effects on P-centre location. This has invariably shown effects of both of these portions on P-centre location; the effect of prevocalic consonant duration on P-centre is very strong (the longer this portion, the later the P-centre); the effects of the post vowel onset to signal offset on P-centre are weaker and appears to vary across subjects, but broadly the longer this section, the later the P-centre. How can the currently proposed determinant of P-centre location be congruent with these findings?

If P-centres are locally determined by the onset of energy around a certain frequency bandwidth (associated with vowel onset), then the effects of prevocalic consonant duration upon P-centre location become highly predictable. The longer this portion of the syllable, the later the P-centre will be. There is thus no difference between the 'global' duration driven models of P-centre location, and the currently proposed model on this parameter.

How can this 'local' model account for the effects of vowel duration on P-centre location? Clearly it cannot. The presented results in this experiment, which state that ramping the offset of a vowel has no significant effect on the settings that subjects provide a degree of evidence that duration of the vowel might not affect P-centre location. If the subjects do not make their settings using information about the offset qualities of a signal, it seems less likely that they might use such information if it consists of duration, rather than amplitude decay information.

This is not impossible however. Contrary to the view implicit in the above paragraphs, subjects might wait until the offset of a signal in order to make a perceptual judgement about where the perceptual centre of that signal lay. A

dynamic rhythm setting task might engender such strategies, by the constant repetition of the signal; the subjects might be able to integrate all the information about the signal and using this to make their settings, or be anticipating the signal's duration when judging the rhythmic qualities of the sequence at any one time.

These issues will be addressed in the next chapter, where the effect of vowel duration upon P-centre location will be tested, with attempts to avoid the artifacts that might appear through the use of natural speech editing or large subject differences.

The conclusions of the current experiment are therefore that the amplitude onset characteristics of a vowel affect the P-centre of that vowel; the same amplitude manipulations to the offset of a vowel do not shift the P-centre to a significant degree. This result is in agreement with the results of Experiments 5 and 6 of this thesis, that increasing the amplitude of a syllable final /t/ burst does not affect the P-centre location of that syllable, and that ramping the onset of a natural vowel affects the P-centre location. The onset amplitude characteristics of a vowel are thus more important than the offset events of a syllable in determining P-centre location.

Chapter Eleven
Experiment 8
The effect of vowel duration on P-centre location

Abstract
The global models of P-centre location (Cooper et al, 1986,1988, Howell 1984, 1988a&b, Marcus 1981) all conceive that the entire syllable duration affects the P-centre of a signal. The local model of P-centre location being devised in this thesis maps P-centres onto characteristics of the vowel onset, and thus the durations of the segments in the syllable do not affect the P-centre *per se*. An explanation has been provided for the effect of prevocalic consonant duration on P-centre location. An explanation is needed for how a local model can account for the effects of post vowel onset duration on P-centre. To resolve this the variation in P-centres caused by altering the duration of a synthetic vowel were investigated, to establish whether there was an effect, and if necessary calibrating the effect for use in the model. The results indicated that the effect of vowel duration on P-centre location, when amplitude profile and spectral content are held constant, is not significant.

11.1 Introduction

From the previous experiments, details have emerged of the relationship between syllabic P-centre location and the acoustic structure of the signal. Amplitude changes in initial/final consonants ("sha" or "t") do not affect the P-centre of a syllable. Intensity changes in the vocalic portion of a syllable do affect the P-centre location. Events at the onset of the vocalic portion are more important than those at the offset of the vowel in determining P-centre location.

The model being developed therefore is taking the rise time/intensity changes of a certain frequency bandwidth (corresponding to attributes of the vowel onset) as a principal determinant of P-centre location. This term will be defined more precisely in the next chapter. In the meanwhile it should be noted that, as

outlined in the previous chapter, such an approach would predict the most consistent experimental finding in the P-centre literature - that of the effect of prevocalic consonant cluster duration on P-centre location, both in production and perception. This portion, as manipulated in previous experiments, does not generally contain the salient frequencies - when the prevocalic consonant cluster duration is manipulated experimentally, low sonority segments such as "sh", "tch", and "t" etc. are generally varied, rather than more sonorous segments such as "m" "w" "l" etc (eg. the experimental manipulations of Cooper, Whalen and Fowler 1986, 1988, Fowler 1979, Marcus 1981, Howell 1984). This prevocalic segment therefore represents a period of "empty time" before the more sonorant acoustic energy begins. The longer this portion, the later the P-centre of the signal.

This approach also predicts the finding that the acoustic vowel onset *per se* is not a perfect determinant of P-centre location, since the model is defined in terms of the perceptual consequences of a critical rate of change of energy around the frequency bandwidth associated with vowel onset, rather than the acoustic onset of this energy.

This approach is defining P-centres as discrete punctate acoustic events which relate to local events within a signal. It is therefore very different to some the models of P-centres described in earlier chapters. Marcus's model is a global one, as is Howell's. Fowler's description of P-centres as relating to vowel articulation is a local explanation, but one which denies the possibility of acoustic modelling. Later work by her and her colleagues has been concerned with the whole signal, however (Cooper Whalen and Fowler 1988). Pompino-Marschall's model is semi-local, in that it describes two perceptual phenomena. From a different perspective, musical beat location has usually been modelled in a local manner (Vos and Rasch 1981, Gordon 1987).

A finding that any hypothesized local model of P-centre location does not predict is the effect of post vocalic onset duration on P-centre location. If P-centres are linked to acoustic attributes of vowel onset, then how could events which could occur hundreds of milliseconds later be also affecting the beat of the sound? This next experiment will address the effect of vowel duration on P-centre location.

The effects of post vowel onset duration has been found to varying degrees by several researchers, although the strength of the effect has varied from researcher to researcher, and in some cases from subject to subject.

11.2 Previous work on vowel duration

11.2.1 Marcus's work

Marcus's original model of P-centre location gave vowel duration a weighting of 0.25 (compared to 0.65 for prevocalic consonant duration). This was based mainly on the P-centres of different spoken digits (1 - 9). This result can be easily explained by the consideration that while these speech items varied in vowel duration, they also varied in vowel identity and quality. Harder to explain is his finding that extending the duration of vowels using pitch-synchronous extension shifts the P-centre of the signals 'back' in time.

11.2.2 Cooper Whalen and Fowler' work

Cooper et al (1986) found that reducing the duration of a vowel shifted the P-centre towards the onset of the signal; but simultaneously with this manipulation they were altering other aspects of the stimulus, namely the gap duration in 'sta'. They shortened the vowel by excising whole pitch periods.

The same researchers (Cooper et al 1988) studied the effect of vowel duration more explicitly in a later experiment. They reduced the duration of a natural /a/ by excising whole pitch periods. They used this vowel series both on its own, and edited into a /sa/ continuum in a type of rhythm setting experiment. The sounds were set against a reference sound /ba/. They found a small effect of vowel duration, but this differed between the /a/ and /sa/ continua, and more importantly, between the subjects. The three subjects showed quite different patterns of interval settings, with one evincing no effect of duration, and the other two showing an effect, but in opposite directions. Cooper et al present this finding as a potentially interesting one, but given the consistency of other P-centre experiments across subjects this must be regarded as a problem in their experiment. In this thesis there have been differences between subjects in aspects of intervals set in dynamic rhythm setting tasks, for example differences in the amount of variation in intervals set (Experiment Seven). Cooper et al's results indicate large and varying differences between subjects on all aspects of adjustment trials.

11.2.3 Fox and Lehiste's work

Fox and Lehiste (1987) in common with results discussed by Howell (1984), found an effect on production timing of vowel duration. As mentioned above such differences could be affected by the vowels having other differences than duration (and also, as they found, some difference between speakers). They found in perception, however, that if the durations of vowels with different qualities were equalized by shortening the vowels then the measured adjustment differences disappeared. This indicated that vowel duration did affect a syllables P-centre.

11.2.4 Seton's work

Seton (1989) varied stimulus duration in a full P-centre determining rhythm setting experiment. He was working with a non-speech saw toothed waveform (400Hz). He used 10 subjects (and fewer trials from each). He found no difference between subjects. The slope of the duration/P-centre relationship was -0.1, as opposed to Marcus's slope of -0.25.

11.3 Experiments which vary post vowel onset duration

Other experiments have been carried out which alter the post-vocalic portion of a syllable by altering the position in time of a syllable final consonant.

Marcus (1981) reported that extending and reducing the period of silence before a syllable final "t" burst in "eight" altered the P-centres of the resulting family of stimuli. This was despite the manipulation being "barely noticeable" (1981 p252). This increased his evidence that P-centres are determined by acoustic events over the whole time course of a signal.

Cooper, Whalen and Fowler (1988), in an extension of their investigation of the effects of vowel duration on P-centre location created two families of "at" stimuli. In the first the vowel was successively shortened by excising pitch periods. In the second the vowel was reduced in length by the same amounts, but the overall durations of the stimuli were kept constant by increasing the duration of the pre "t" burst silence by equivalent amounts. They predicted that the P-centres of the first, duration varying stimuli should shift with duration. The P-centres of the second, duration constant set of stimuli should not change. The intervals set with these stimuli against a reference sound ("ba") were analyzed. Again their was considerable inter-subject variation, in range of settings made and in effects of vowel duration and overall duration. They

concluded that their results supported the view that "the components of the syllables rhyme have equal or near equal effects on P-centre location" (1988 p30).

11.4 Aims of this experiment

There is thus a body of evidence which indicates that the duration of a syllable's vowel/rhyme duration affects the P-centre of the syllable, albeit in a weak way. In several cases this finding is contaminated by a consideration of how else the vowels might have differed than just duration (ie. the vowels in 'eight' and 'nine' vary in duration, but also in spectral content and amplitude profile). In several experiments however the vowel had been edited to be longer or shorter and this has affected the P-centre.

Natural speech sounds vary a great deal over time. Their temporal structure is vital to their phonetic identity and speech-like qualities. Therefore, even if pitch periods are added or deleted from a 'steady state' of a vowel, it cannot be guaranteed that this manipulation will not disrupt the 'naturalness' of the sound. It is difficult to edit a natural speech sound in this way and alter the duration of the sound only (Seton 1989). Even if the syllable is still perceived as natural, other aspects of that syllable will have been affected, and this could be sufficient to affect the setting of this sound to a rhythm. In this current experiment a synthesized vowel (the Haskins Lab. /a/ used and described in the previous experiment) was be edited to form a continuum of five different durations. The original vowel was edited from a "ba" syllable (described in Kuhl and Miller 1978), and there is no difference across the whole temporal extent of the signal in its spectral content. The effect of altering the duration of this vowel on the P-centre will be measured. The effect of duration on P-centre, when nothing else about the stimulus has been altered, will be determined. The hypothesis is that if the vowel duration alone affects the P-centre of the signal,

then there will be a systematic relationship between the duration of the signal and its P-centre.

The alternative hypothesis is that the vowel duration will have no effect on the P-centre of the signal, since it is the onset characteristics of the syllable which determine the P-centre location.

11.5 Method

11.5.1 Stimuli

A stimulus continuum varying in duration only was created by editing a synthetic vowel (described fully in Experiment 8). The fundamental frequency of the vowel was steady at 114Hz. The onset rise time was held constant at 5ms, and the duration was edited to 76ms, 130ms, 176ms, 230ms and 280ms by excising sections from the offset. The duration increments are not always equal to prevent a signal ending mid-pitch period. These were used, along with the reference sound, in a dynamic rhythm setting task where each sound was set against itself and every other sound. The A - A interval was 1250ms. The stimuli were presented via ER2 insert earphones at 1V peak to peak amplitude.

11.5.2 Subjects

Two experienced subjects (SKS and DC) took part in the experiment. Subject DC was naive to the aims of the experiment; subject SKS was not naive to the experimental manipulation.

11.5.3 Procedure

The five stimuli and the reference sound were used in a full dynamic rhythm setting task to determine the P-centres of the stimuli. There were thus 6 stimuli, and 6X6 experimental stimuli combinations. Each subject performed 8 trials in each stimuli combination.

The trials where the same sound is set against itself, and which normally act as a control, were not included in this experiment as the subjects were both experienced and reliable, missing these trials out shortened the experiment by 48 trials for each subject. Each subject thus performed 240 individual trials.

11.6 Results

stimuli	280ms	230ms	176ms	130ms	76ms	ref
280ms		624.34 13.88	639.12 16.32	633.38 22.10	630.62 18.18	623.68 23.56
230ms	622.81 16.40		640.25 19.69	633.31 17.16	625.81 10.23	633.59 23.52
176ms	615.50 22.75	623.81 12.20		626.75 17.10	623.00 18.12	640.45 20.68
130ms	609.75 22.64	612.75 19.56	615.50 19.63		629.06 16.97	625.32 21.34
76ms	615.94 18.38	610.50 36.30	623.00 19.22	623.06 14.46		627.82 14.78
ref	618.00 19.40	630.64 21.92	613.09 22.40	625.55 22.46	619.68 20.41	

Table 11.1 of means and standard deviations of interval settings of combinations of stimuli which vary in duration (ms). Shaded blocks indicate combinations not performed.

Table 11.1 above shows the means and the standard deviations of the final interval settings. When these values deviate from 625ms, that is from physical isochrony, then this indicates that the stimuli concerned have differing P-centres. The means do not appear to deviate widely from isochrony, although there are differences. Whether these are systematic will be seen when the P-centres are calculated. The standard deviations are quite large in places. Could this be due to the two subjects making different settings?

stimuli	280ms	230ms	176ms	130ms	76ms	ref
280ms		9.063 10.279	18.750 10.201	19.000 13.342	13.625 12.889	20.125 13.55
230ms	14.312 7.445		18.500 16.448	13.563 13.120	7.562 6.663	19.875 15.196
176ms	20.125 13.421	8.812 8.216		14.375 8.679	13.250 12.053	19.625 13.995
130ms	20.125 18.129	19.750 11.246	16.125 14.287		12.687 11.568	19.125 13.672
76ms	16.688 11.324	31.750 20.091	16.250 9.588	10.687 9.555		13.375 7.256
ref	16.562 10.145	13.312 12.632	23.375 13.266	16.938 16.886	14.437 14.00	

Table 11.2 means and standard deviations of absolute deviations from isochrony for all stimuli combinations of stimuli which vary in duration only (ms). Shaded cells indicate combinations not performed

Table 11.2 shows the means and standard deviations of the absolute deviations from isochrony for each stimuli combination. This gives a measure of the absolute amount by which a combination of stimuli had to be adjusted from perfect isochrony in order to acheive perceptual regularity. The larger the mean deviation, the greater the absolute deviation from isochrony.

The significance of the stimuli combinations on the deviation from isochrony was tested by regressing the deviations against the stimuli combinations.

Subjects were included as a predictor in the regression to test whether, as has been previously found, there are large differences between subjects in the effects of duration on their P-centre settings.

The standard deviations shown in Table 11.2 are in some cases larger than the means - therefore the data was transformed using a square root function. This reduced the overall mean and standard deviation from 16.131 (12.882)ms to 3.6637 (1.647).

The significance of the predictors **stimuli combination** on the transformed deviations was tested using multiple linear regression, fitting the equation:

$$y = c + \alpha(x_1)$$

where y=deviation from isochrony, x1=stimuli combination

The regression had the equation:

transformed deviation from isochrony = 3.90 - 0.0106(stimuli combination)

The predictor **stimuli combination** was not significant ($p > 0.05$).
To test for subject differences, the predictor **subjects** was entered as a dummy variable (ie. variable $\beta(x2)$ entered into the above model instead of x1); this was not significant ($p>0.05$).

Thus the different stimuli combinations did not affect the set intervals significantly, and neither did the subjects. The subjects are consistent with each other in the interval settings they made; and the experimental manipulation of duration variation did not have a significant effect on their settings.

11.7 P-centre calculation

The lack of significance of vowel duration on P-centre location means that this manipulation is not one which has a large effect on P-centre location (and thus subjects' settings). The P-centres might still vary with the stimuli durations to some degree, so the P-centres were calculated for the stimuli using the P-centre algorithm. They were all then adjusted such that the P-centre of the reference sound was equal to zero.

The differences in P-centre are small, but seem to be consistent in their variation. The P-centre shifts away from the onset of the stimulus (becomes more negative) as the stimulus duration increases.

Stimulus	Duration (ms)	P-centre (ms)
1	280	-8.00
2	230	-7.66
3	176	-3.66
4	130	0.67
5	76	-1.00

Table 11.3 of Stimulus durations and calculated P-centres (relative to reference sound P-centre = zero)

To quantify the significance of this relationship, the P-centres were regressed against the predictor **stimulus duration** with linear regression, fitting the equation:

$$y = c + \alpha x$$

where y=P-centre and x=duration

The regression has the equation:

P-centre = 3.87 - 0.0437 duration

The predictor **stimulus duration** is significant ($t_{1,3}$= -3.969, $p < 0.05$).

Figure 11.2 shows the relationship plotted graphically. The longer the vowel duration, the later the P-centre of the syllable. The effect is small; over a change in vowel duration of 200ms, there is a shift in P-centre of 7ms.

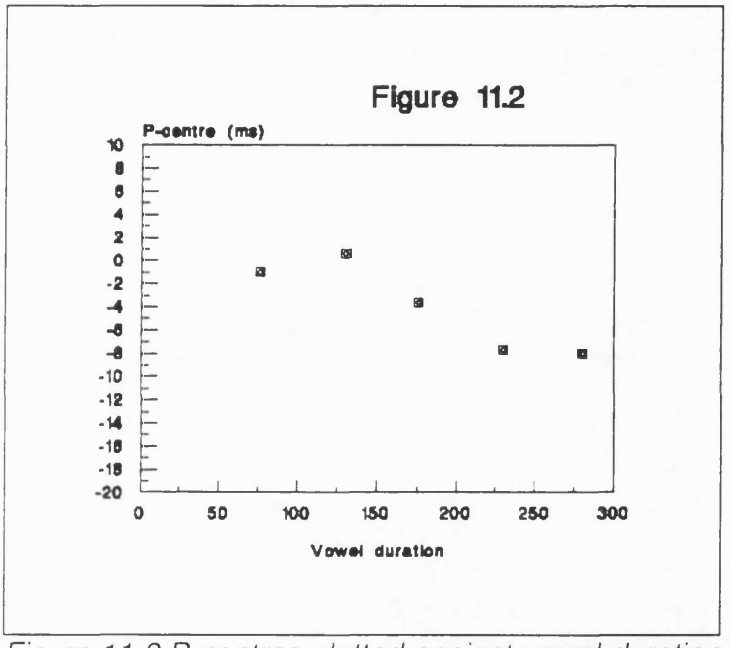


Figure 11.2 P-centres plotted against vowel duration

The plot of P-centres against stimulus duration is roughly linear - there is a suggestion that at the longer vowel durations the graph is beginning to 'level off', perhaps at even longer stimulus durations the graph would take on an exponential appearance.

The observed variance in the P-centres of the signals is thus accounted for by the duration of the stimuli. The relationship between P-centre and duration,

whilst significant, has a very shallow slope (-0.0477) as opposed to that previously noted (in this thesis) for rise time and P-centres (-0.235).

11.8 Other P-centre model predictions

11.8.1 Marcus's model

Marcus's model assigns a reduced weighting to the effect of vowel duration on P-centre location. The relationship between P-centres and the vowel duration has the slope 0.25. Marcus's model was therefore used to make predictions about the P-centres of the duration varying vowel stimuli, to compare against the observed P-centre variation. These are shown in Table 11.4 below.

Duration (ms)	Marcus Model P-centres (ms)
280	-70.0
230	-57.5
176	-44.0
130	-32.5
76	-19.0

Table 11.4 Marcus model P-centre predictions for stimuli

11.8.2 Howell's model

Howell's model predicts a linear effect of vowel duration on P-centre location, with a slope of 0.5 if all other factors are held equal (such as stimulus rise time, amplitude).

Using the software described in Experiment three, predictions of the centres of gravity of the stimuli were calculated, and are shown in the Table 11.5 below.

Duration (ms)	Howell model predictions (ms)
280	-148.8
230	-122.3
176	-95.4
130	-71.2
76	-44.0

Table 11.5 Howell model predictions of P-centres for stimuli

11.8.3 Cooper et al's predictions

Although their perspective on P-centres would preclude their formulating a model *per se*, Cooper et al (1988) do provide information as to the effect of manipulating vowel duration on the amount of offset needed to lead to a perceptually even rhythm (and hence a difference in P-centre). As outlined in the introduction however, this experiment contains such a range of differences as to the effect vowel duration has on rhythm settings as to render the presented results unusable for hypothesis testing. Indeed they state that this difference is due to individual differences without outlining what might have caused such differences. This is vital, to avoid criticism about the general uniformity of subjects responses in P-centre experiments.

In the experiments described in this thesis, there have twice been significant subject differences; once due to one subject making unusual settings with certain stimuli combinations; once due to a large degree of variation from one of the subjects (the author). There has nowhere been the very large, consistent difference between subjects that Cooper et al (1988) report.

Thus a test cannot be made of whether these results confirm Cooper et al's (1988) data; one of their subjects did show no significant effect of duration on the intervals set in a dynamic rhythm task, and these results are similar to those. The finding of no significant difference between the subjects in this experiment indicates that Cooper et al's assertion that vowel duration/syllable rhyme has a differential effect on P-centre location, according to the listener, was not replicated.

11.8.4 Vos and Rasch predictions

Vos and Rasch (1981) developed a model of perceived onset time which corresponded to the P-centre of the signal, and which used the parameter of intensity change at the onset of the signal. This model, and any local model of P-centre location, would make the prediction that increasing the duration of the vowel would have no effect on P-centre location. This model has previously not predicted the results of speech manipulations well; it is included here to provide an example of a model which does not utilize information about duration. The software described in Experiment three was used to provide Vos and Rasch model predictions; these are shown in Table 11.6 below. They do not vary across the different stimuli durations.

Duration (ms)	Vos and Rasch Predictions (ms)
280	-6.2
230	-6.2
176	-6.2
130	-6.2
76	-6.2

Table 11.6 Vos and Rasch model predictions for stimuli of different durations

The predictions of the Vos and Rasch, Marcus and Howell models are all shown plotted against stimulus duration in Figure 11.3, in addition to the experimental P-centres.

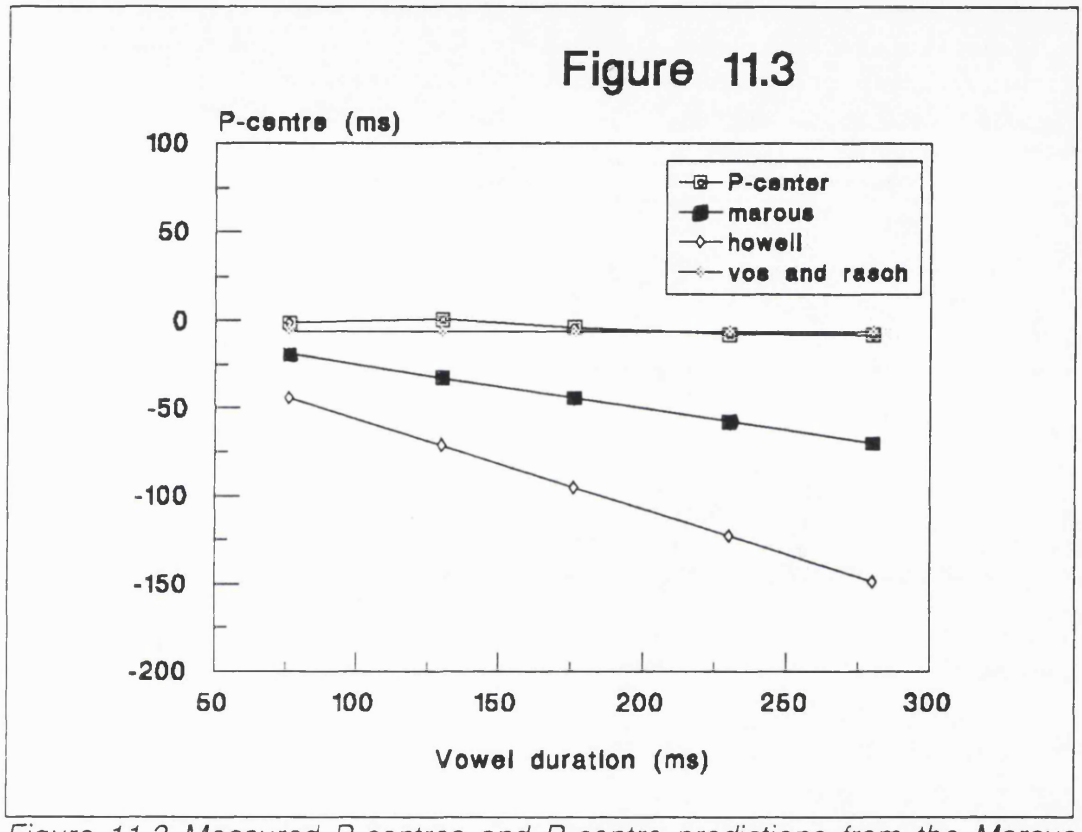


Figure 11.3 Measured P-centres and P-centre predictions from the Marcus Howell and Vos and Rasch models plotted against vowel duration

Which model of P-centre location best predicts the observed variance? Figure 11.3 shows the predictions of the Howell and Marcus model against duration, along with the observed P-centres. It can be seen that, although there is variation across the P-centres of the vowels with different durations, the relationship has a much shallower slope than that predicted by either Marcus's or Howell's models. The predictions from the Vos and Rasch model of perceived onset of musical tones provides a much better fit, despite this model's predictions not varying across stimuli.

11.9 Discussion

If a synthetic vowel is edited to produce stimuli with different durations but identical onsets, spectral contents and temporal spectral changes, the duration of the vowel affects P-centre. The amount of P-centre shift caused by this manipulation is however very small. There was P-centre shift of around 8ms for a 200ms difference in vowel duration. The regression of P-centre against vowel duration was significant (r^2(adj) = 75.9%, p = 0.035), but the regression of absolute deviations from isochrony in the rhythm setting task on the stimuli combinations was **not** significant (r^2(adj) = 0.0%, t = -1.20, p = 0.235). This implies that whatever P-centre shift is caused by the duration manipulation, it is not large enough to affect the subjects' settings significantly. This casts doubt on the importance of vowel duration on P-centre location.

11.10 Relationship of this result with previous work

The results of this experiment thus indicate that if spectral contents are held constant, vowel duration has only a slight effect on P-centre location. This results appears very different from the work outlined in the Introduction section, which described experiments which had found varying effects of vowel duration on P-centres. However the results do fall within the range of previous experimentation; this section will address the similarities and consider reasons for the differences which remain.

First the possibility that no significant effect of vowel duration was found because entire experiment failed must be addressed. There are two reasons why this is unlikely. As mentioned above this results falls within the existing range of the effects of vowel duration. In addition the subjects do not differ

significantly from one another in the settings they make; they have taken part in previous rhythm setting experiments; one of the subjects was naive.

11.10.1 Direction of P-centre shift In common with previous research the shift of P-centres with vowel duration is away from the onset of the vowel as duration increases. Longer vowels thus lead to later P-centres. This direction of shift is the same as that found by other researchers (Marcus 1981, Cooper et al. 1988, Fox and Lehiste 1987).

11.10.2 The effect of vowel duration is within the range of previous experiments The regression of subjects' set intervals against stimuli combinations was not significant, indicating that the vowel duration was not affecting their rhythm setting judgements. When the P-centres were calculated from these intervals the observed **shift** in the P-centres was small but it was significantly accounted for by the vowel duration. As mentioned above this slight effect of vowel duration has been noted by Cooper et al (1988) in one subject's responses. The slope of the P-centre/duration relationship is of a similar order to that noted by Seton (1989), (the P-centre/duration slope is -0.0437 in this experiment, -0.1 in Seton's). As is shown in Figure 11.3, the relationship between P-centre and vowel duration observed in this experiment is much weaker than that predicted by Howell or Marcus, and is more similar to that of Vos and Rasch (which predicts no effect of duration on P-centre).

The small effect of vowel duration shown in this experiment thus falls within the range of previous vowel duration effects which have not indicated a single concrete effect of vowel duration on P-centre location.

This is not to avoid the fact that vowel duration effects have been seen previously. How can the difference between the results be explained?

11.10.3 Experimental Task

This experiment involved a full rhythm setting paradigm to establish P-centres for the stimuli - only Marcus (1981) also carried out full P-centre setting tasks. Others used a reduced paradigm (Cooper et al 1988, Fox and Lehiste 1987, see Chapter One), which allows for fewer degrees of freedom in the measurements. This may lead to larger effects being found than remain after more exhaustive testing.

If the differences are not due to varying task requirements, then the issue of uncontrolled variables affecting the experiments must be addressed.

11.10.4 Rhythmic illusions Lame (1990) in his thesis was working on time perception and control; he was interested in temporal illusions. Thus the basic assumption of all P-centre experiments - that the intervals between stimuli do not affect the perception of the stimuli and vice versa - was of interest. Whilst accepting the assumption that signals have a perceptual moment of occurrence, which is context free, the contention that in a perceptually isochronous sequence, the P-centres are physically isochronous, is rejected. Instead, Lame stated that the P-centres are only "perceptually isochronous, and given the prevalence of ... temporal illusions... that is no isochrony at all" (1990, p180). Lame provides two possible hypotheses to account for the effects of vowel/rhyme duration on P-centres as measured in dynamic rhythm setting experiments.

The first is that an internal clock is reset to zero at a fixed point in the syllable (ie. syllable or vowel onset). From this perspective, the vowel length will affect the dynamic rhythm setting task by causing temporal distortion. In a perceptually regular sequence, the P-centres are not physically regular.

The second hypothesis is that there is no precise moment of occurrence, instead different aspects of the syllable affect the internal clock to different amounts. This hypothesis predicts no effect of temporal distortion due to vowel/rhyme duration, due to a cancelling of the induced phase shift by a change in the phase of entrainment.

Do these results support or reject Lame's hypothesis? The argument becomes rather circular; A rhythm setting task is used to determine P-centres and a parameter of the stimuli (vowel duration) is implicated in P-centre location as measured in this way. This same parameter is then found to affect how subject's perform the rhythm setting task upon which the P-centre results are based. If Lame is correct, then one of the assumptions of the P-centre determining paradigm - that is a lack of context sensitivity - is incorrect. A solution to this problem cannot be determined from the current or previous experiment, although the disappearance of any profound effect of vowel duration on P-centre location when simple synthesized speech is used provides some circumstantial evidence for Lame's position.

11.10.5.2 Vowel structure This experiment was designed such that only the duration of the vowel was varied. This was controlled for by using a synthetic vowel with a constant pitch. The duration of this synthetic vowel did not significantly affect the subject's settings. As discussed in the Introduction previous researchers all used edited natural vowels (Cooper et al 1988, Fox and Lehiste 1987, Marcus 1981). The formants which characterize the qualities of natural vowels are relevant both in their frequency range and their patterns of change (formant transitions). It is possible therefore that vowel duration *per se* is not the important attribute that is affecting the P-centre judgments set in experiments which use edited natural vowels. Instead the vowel contents might be better characterized in terms of the formant transitions.

11.11 Implications for modelling

The role of post vowel onset attributes in P-centre location has not been entirely clarified by these results. Instead vowel duration itself, if all other factors are constant, has been shown to not significantly affect P-centre judgements. Reasons for the previous vowel effects were considered in terms of temporal illusions and the temporal structure of vowels; this is speculation and further experimentation will be needed to provide a comprehensive explanation. This does not however provide a clear guide for modelling. In the modelling chapter of this thesis therefore P-centres will be determined as local frequency specific onset events; these are the attributes the previous experiments have shown to affect subjects' P-centre judgements. The inclusion of other acoustic information - such as post vowel onset attributes - is not precluded by such an approach.

11.12 Conclusions

In conclusion therefore these results indicate that:

1) Vowel duration manipulation does not have a significant effect on the absolute deviations from isochrony in the settings that subjects make in a dynamic rhythm setting task.

2) The P-centres of these stimuli do vary slightly with vowel duration; the relationship is a weak one (slope = -0.0437), especially when compared to the predictions of other models (Howell 1988, Marcus 1981). The relationship most closely resembles the predictions made by the Vos and Rasch model of perceived onset time. This model makes the prediction that the P-centre would not vary at all with vowel duration.

3) There was no evidence to support Cooper et al's statement that there are large differences between listeners in terms of the effect of vowel duration on P-centre location. This result is negative in that there was a failure to find a difference, but there was no indication of the large differences Cooper et al

found between just three individuals. They did not provide any criteria for discriminating between those who display such a difference, and therefore none could be applied to the subjects. Repeating the entire experiment would provide further evidence on this point; time meant that this was not a possible option.

These results indicate that a local model of P-centre location, based around the intensity change characteristics of the rise of certain frequencies would be valid. In the next chapter the spectral and intensity attributes of a signal at onset, as determinants of P-centres, will be quantified. A qualifier is that these are not necessarily the only parameters which play a role in the perception of P-centres.

Chapter Twelve
A local model of P-centre location

Abstract
A conceptual framework for modelling P-centres in a local manner is outlined. A protocol and a method for the modelling is described, as well as the reasoning behind every stage of the process. The model is tested against all the experimental stimuli for which P-centres have been determined in this thesis. The resultant model of P-centre location is discussed in terms of previous models of P-centre location.

12.1 Introduction - how should P-centres be modelled?

The experiments described in this thesis have indicated that the rise time of a signal affects its P-centre, if the onset of the signal has certain acoustic characteristics. Thus the onset qualities of a syllable affect the P-centre if the onset segment is of high sonority (a vowel or semi-vowel - see Chapter Nine) rather than low sonority segment (such as a high frequency fricative). The relevant acoustic elements broadly correspond to those of the vocalic portion of a syllable. In this chapter an attempt will be made to characterise attributes of the signal onset, and see if P-centres can be mapped onto this value.

The hypothesis therefore is that onset of the vowel, or vowel like sounds in a syllable lead to the percept of a rhythmic centre. An attempt will be made in this chapter to attribute P-centres to frequency dependent amplitude changes within a syllable. Since this thesis has been guided by a concept of P-centres as parallel phenomena to 'beats' in non speech sounds (Howell 1984, Morton et al 1976), and P-centres are being modelled in a local manner, after the musical beat location literature (Vos and Rasch 1981, Gordon 1987), the onset attributes that lead to P-centres will not be defined as vocalic, although they will

be broadly similar to qualities of the vowel onset. In other words, P-centres will be modelled in terms of the **onset** of the **increase in energy** of a **certain frequency bandwidth** within the syllable.

12.2 A modelling protocol

P-centres are thus to be modelled as local syllable events, not as dependent upon the entire syllable structure.

This approach is predicated upon there **being** some particular frequency bandwidth, changes in the onsets of which correlate with the P-centres. The protocol for determining this bandwidth was thus that:

i) A set of speech items, the P-centres of which have been established, would be selected

ii) These speech items would be broken down into *n* frequency components, using a filterbank.

iii) The *n* channel outputs from the filterbank, for each speech item, would be analyzed in terms of the onset characteristic *r* - for example, rise time.

iv) The value of *r* for each of the *n* channel outputs for each of the speech items would be compared with the P-centres for each of the speech items.

v) If the values of *r* for any one of the *n* channel outputs, *r'*, corresponds well with the P-centres of all the speech items, then this value will be calculated for a test set of different speech

items, to examine whether it makes good predictions about these P-centres. If the values of *r'* are a good fit for the P-centres of the different set of speech sounds, the model will be implemented computationally.

The modelling protocol is shown in Figure 12.1. The analysis tools, stimuli and parameters to fit each level of the protocol are described in the next section.

12.3 Modelling P-centres - the method

12.3.1 the speech items

Rather than use some of the experimentally controlled speech items, such as the onset ramped stimuli from Experiment 6, it was decided to attempt to model P-centres using a set of natural speech items, which vary unsystematically. This was to avoid systematic variation in the model being a function of controlled parameters in the original speech.

The collection of speech items used was the "one"s and "two"s for which P-centres were determined in Experiment 3. This corpus has the disadvantage of containing only two word types, but the strength of containing items which have very different acoustic characteristics (which can be seen to some extent on the oscillograms - Figure 5.2). A model of P-centres based on these would not contain implicit assumptions about 'normal' British English speech; the corpus contains items from speakers from Lancashire, Lincolnshire, Glasgow, Belfast, Nottingham, London and Winchester. In addition both female and male speakers were represented (5 female, 3 male).

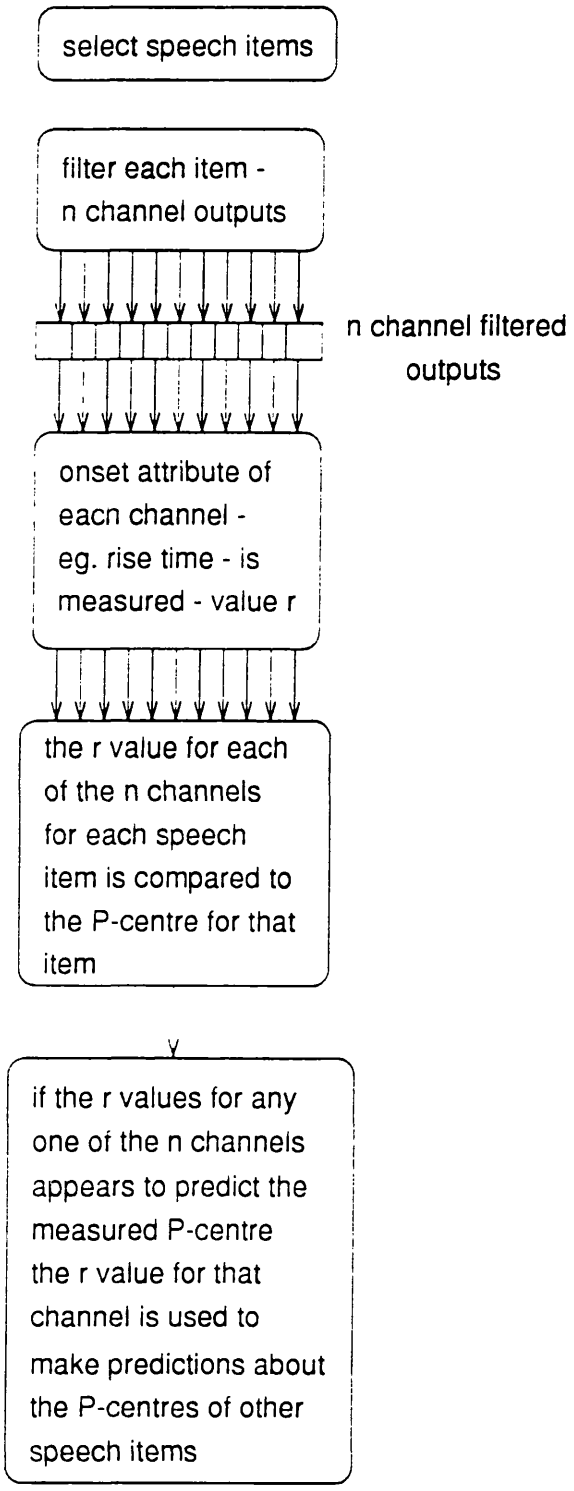


Figure 12.1 P-centre modelling protocol

12.3.2 The filterbank

Since this model might be argued to be approaching a perceptual, rather than a purely acoustic level of analysis, it was decided to use a perceptually plausible filterbank. One method for modelling the frequency processing attributes of the peripheral auditory system is an implementation of a bank of linear bandpass filters. The filterbank used in this modelling of P-centres is a GammaTone filterbank. The GammaTone Filter (GFT) is based upon the technique of 'reverse correlation' (de Boer and Kuyper 1968). This technique is a method for comparing the responses of a primary auditory nerve fibre to white noise stimuli to the original input. This is used to measure the auditory filter shape. The impulse response of the auditory filter, that precedes the spike generation mechanism of that filter, is broadly represented by the 'revcor' function. The GTF is an analytic mathematical function that approximates measured revcor functions (Johannesma 1972). The form of the GTF affords filter properties to be derived analytically. Since the GTF is derived from measured impulse responses, it has complete amplitude and phase information and not just amplitude information as is the case with filters derived from psychoacoustic masking experiments (eg. the roex filter, Patterson and Moore 1986).

Holdsworth, Nimmo-Smith, Patterson and Rice (1988) presented a digital multiple pass infinite impulse response (IIR) filter scheme for the implementation of the GTF. The implementation used in this thesis was written by Mark Huckvale at the UCL Phonetics Dept., with additions made by Richard Baker, using the original "C" software of John Holdworth. The UCL Phonetics implementation was designed to provide the frequency analysis component in a computer model of speech perception (Darling 1991). It thus had the twin advantages of availability and perceptual plausibility. Figure 12.2 shows the output of this filterbank on the word "one", displayed as a set of oscillograms.

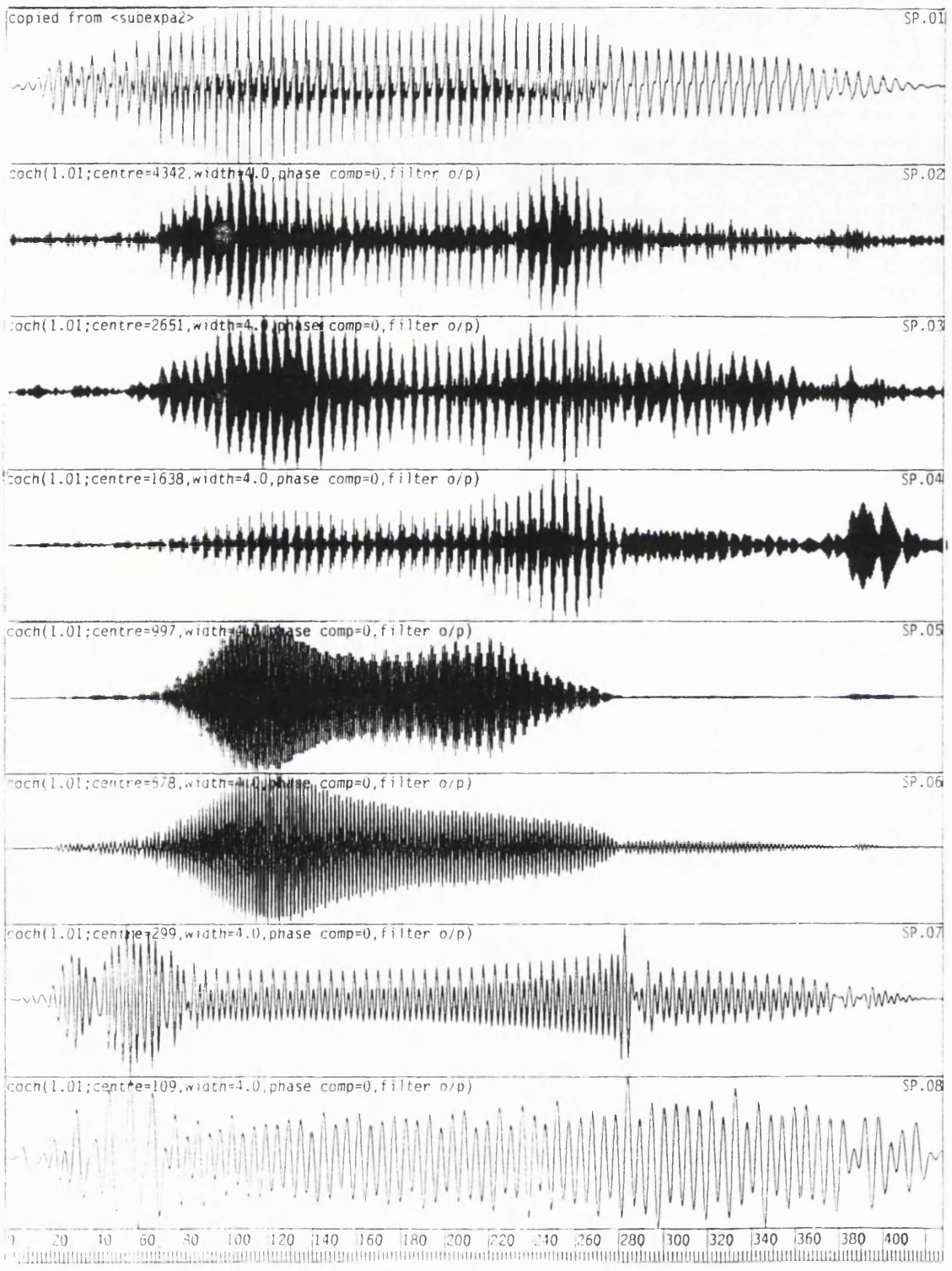


Figure 12.2 Oscillogram of word "one", and the output of the APU GammaTone Filterbank when this signal is passed through (number of channels=7, ERB=4.0).

12.3.3 The onset parameter

So far in the thesis, the term 'onset events' has been very generally used, without any formal operationalization of the term. The parameter to be measured is rise time; this has been shown to affect perceptual phenomena, for example temporal order judgements, onset categorization musical note perception (see Chapter Two for further details). This is defined acoustically as the time the signal takes to pass between 10% and 90% of the maximum amplitude. The larger the rise time value, the longer the signal takes to reach maximum amplitude, and the 'slower' the onset of the sound. The use of rise time has the advantages of simplicity and ease of measurement; it has the disadvantage of assuming linearity in the amplitude onset, which may be non-linear in character. The linearity/non-linearity of the amplitude onset need not yet be addressed in the modelling; it is important to first establish that the general attributes of the onset amplitude change itself can be used to predict the P-centre location. The rise time provides a rough measure of the rate of change of amplitude at the onset which is useful for this task.

In this analysis, the measurement of rise time in *n* frequency channels, it is likely that the amplitude profiles of some channels, for some speech sounds, will not occur at or near to the physical onset of the overall signal. Thus in "two" the lower frequency contents which make up the vowel portion do not begin to increase until after the high frequency noisy "t" burst at the onset. In addition to a parameter containing information about the **slope** of the onset of a channel output, information is needed about **where** in time this rise time occurs, given that it need not be the physical onset of the whole unfiltered signal. Thus a value of where in relation to the start of the whole signal 50% of the maximum amplitude is passed was used. The parameters are shown in Figure 12.3 on a schematic amplitude envelope.

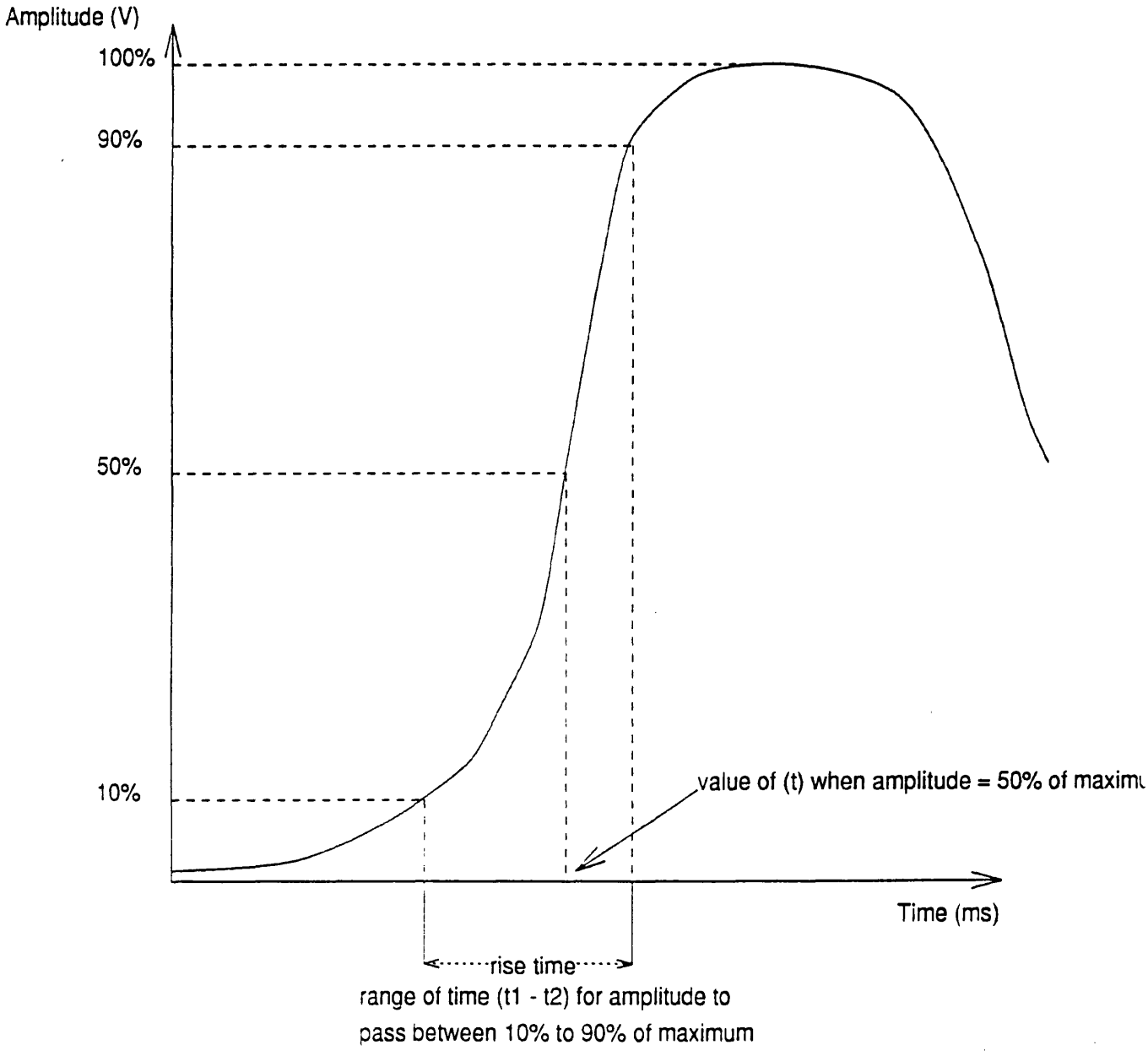


Figure 12.3 Schematic amplitude envelope plot showing the parameters rise time and 50%_max.amp

Thus the two onset parameters used - the rise time and the $50\%_{max\_amp}$ values were chosen. In reality there is likely to be some correlation between these two parameters in certain cases. For a spoken work "one" articulated with a long period of lip spreading (that is, a long "w" sound) and a gentle onset, the rise time and the $50\%_{max\_amp}$ values for the entire signal would correlate. The longer the rise time, the later the $50\%_{max\_amp}$ value. In other cases there will be no correlation; in the spoken word "two" the frequencies relating to the vowel onset may increase sharply at vowel onset (small rise time value), while a long "t" burst may lead to the vowel onset occurring late in the syllable, leading to a large $50\%_{max\_amp}$ value.

In terms of the patterns of energy in the *n* channel outputs, these two parameters represent where there is a rapid increase in amplitude, and the qualities of that increase (fast or slow).

12.3.4 The analysis

Two methods of analysis were used. The first was **exploratory**, and involved using 3-D plots of the data to highlight any clear associations between P-centres, onset parameters and frequency channel.

The second, statistical, analysis was guided by the observational examination. This was in the form of correlations to eliminate correlating channel outputs, and stepwise regression to establish the best predictors for the P-centre values.

12.4 Modelling P-centres - the procedure

The set of sixteen speech items (the "one"s and "two"s from experiment three) were processed in the following manner - shown on Figure 12.4.

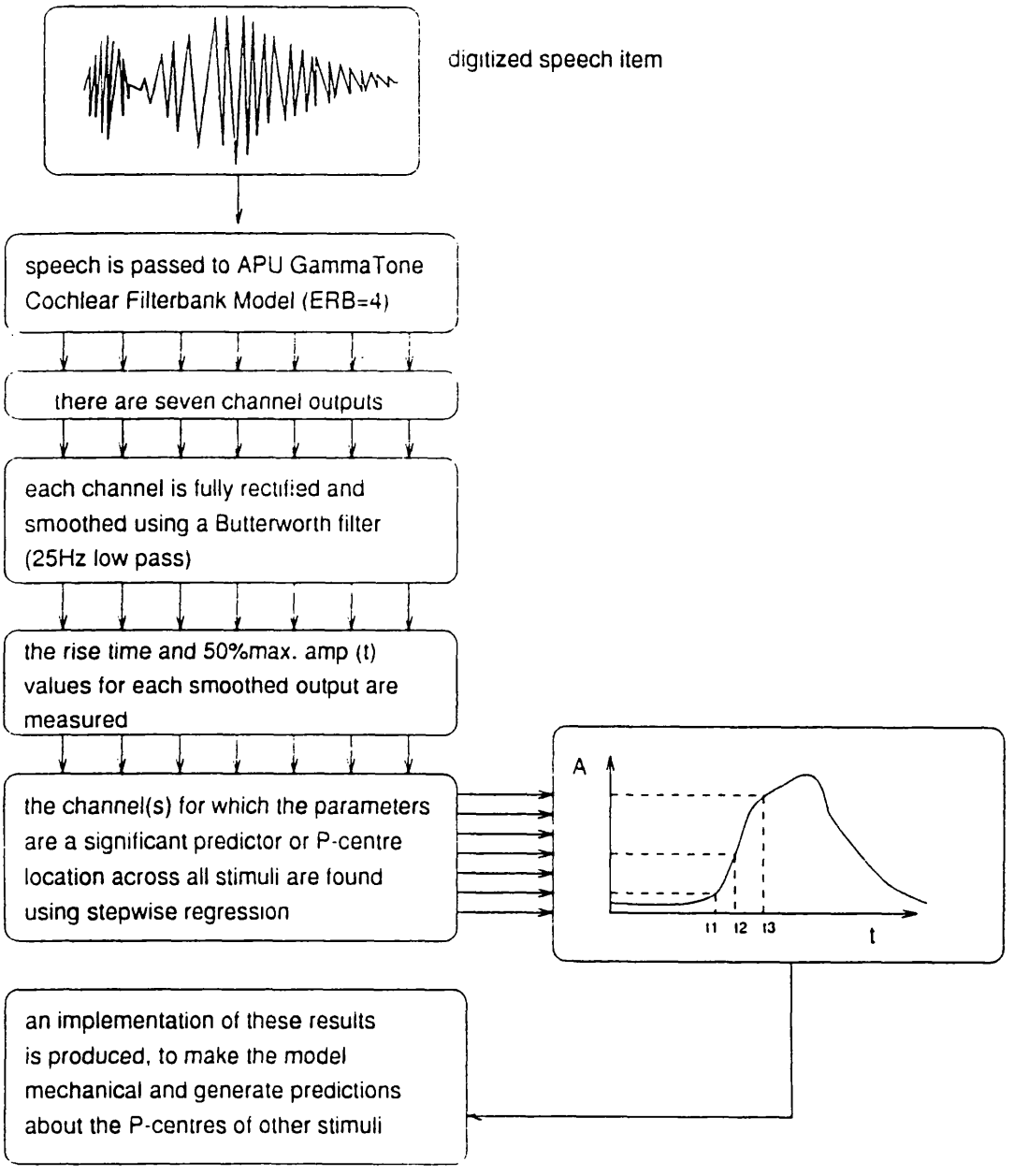


Figure 12.4 Diagram of process for modelling P-centres.

▸ The speech item was filtered using the implementation of the APU GammaTone filterbank.

▸ Each channel output was fully rectified, then smoothed using a Butterworth filter at 25Hz.

▸ The rise time of each smooth rectified channel output was measured using an SML program, and the $50\%_{max\_amp}$ point measured.

All of this processing was digital, carried out on the SFS speech files themselves; the process was controlled by a C-shell script on the MASSCOMP 5.0 RTU machine. The code for the rise time program, the details of the rectifying and smoothing programs, and the C-shell script are given in the APPENDIX.

12.5 Modelling P-centres - the results

12.5.1 Data description Figure 12.5 shows a plot of **rise time** (y-axis) against **P-centre** (z-axis) against **channel centre frequency** $\log_n$ (x-axis). The filterbank was set to n = 27 channels (centre frequencies = 4938, 4342, 3828, 3382, 2993, 2651, 2350, 2084, 1848, 1638, 1450, 1283, 1132, 997, 876, 766, 668, 578, 498, 425, 359, 299, 244, 195, 150, 109, 72 Hz) spaced at 0.5 ERB; plotted in $\log_n$. Therefore for each of the sixteen speech items, the rise time for each of the 27 channels is plotted, leading to 432 data points. The surface fitted to this (r^2 = 0.33948) represents the best 3-d fit to these data points, and can be used to identify patterns in the data points. The surface shows a prominent ridge along the 'nearest' side, which slants upwards from right to left. This indicates a relationship between lower centre frequency channels, rise times and P-centres - at the lower end of the frequency axis, P-centres are later (more negative) as rise times increase (are more positive). Thus P-centres appear to vary with rise times, at lower frequencies - **longer rise times** at **lower frequencies** are leading to **later P-centres**. This figure plotting is simply

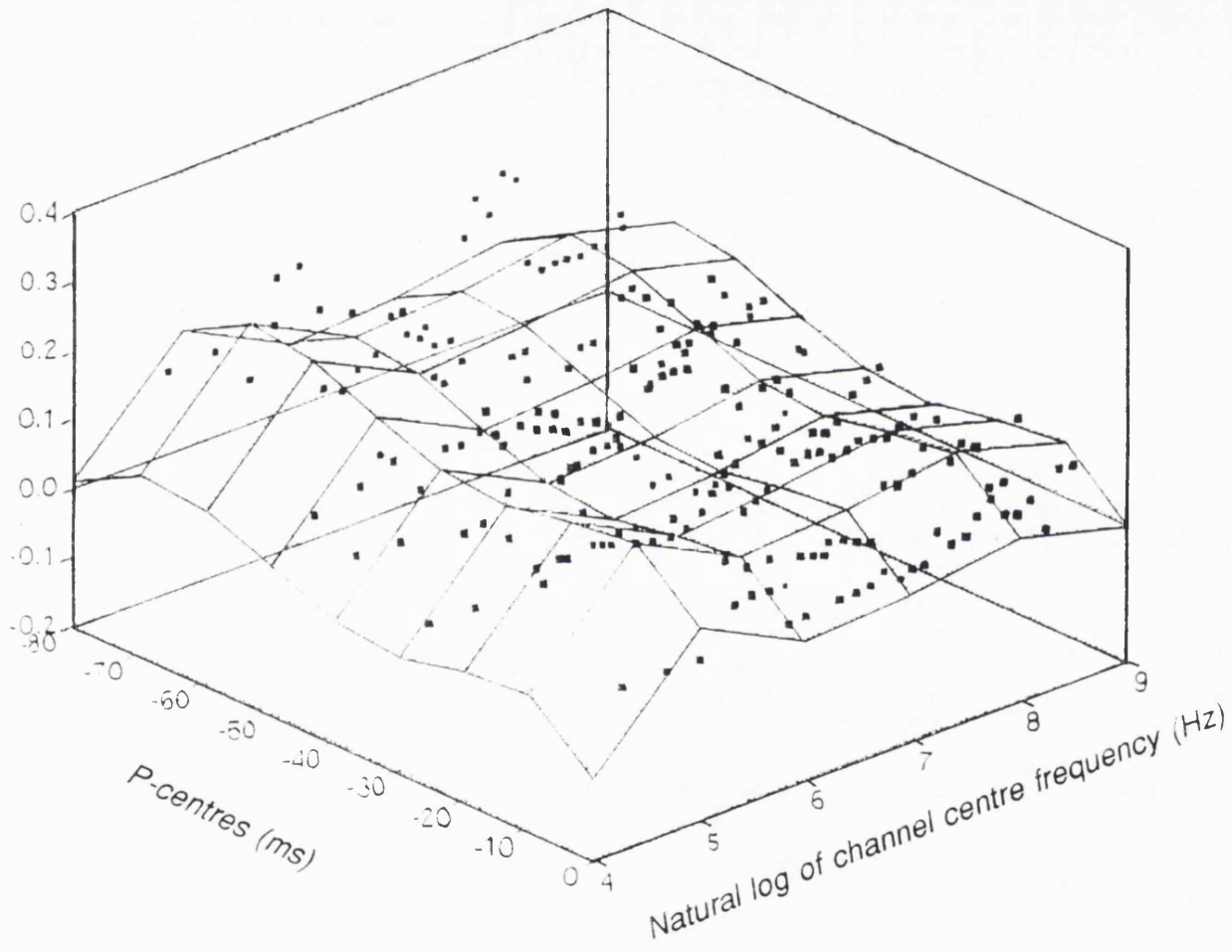


Figure 12.5 3-D plot of P-centres (z-axis) against $\log_n$ centre frequency of filterbank channel (x-axis) and channel signal rise time (y-axis). A surface has been fitted to the plot; note the 'ridge' on the near side of the surface indicating a relationship between P-centres and rise times at the ***lower*** *frequencies.*

a method of describing the data, and represents a guide for more analytical techniques to detail the relationships between variables.

12.5.2 Statistical analysis This analysis was performed with the channel width of the APU GammaTone filterbank increased 4.0 ERB (as opposed to the 27 channel analysis, with the channel width 0.5 ERB). This gave 7 channel outputs with centre frequencies of 4342, 2651, 1638, 997, 578, 299 and 108 Hz (this was to allow regressions to be carried out - 27 predictors are too many for 16 speech items). Figure 12.6 shows these different channels for the word "two" from speaker SR. It was hypothesized that this output would provide a broad frequency breakdown image in which an overall pattern of relevant frequency information would emerge, rather than a high resolution image of the frequency information, which an analysis with more channels would provide.

The P-centres of the speech stimuli, and the rise times and the $50\%_{max\_amp}$ values for all the channel outputs were analyzed statistically to determine which, if any, of the channel outputs varied with the P-centres of the stimuli.

The first analysis was a STEPWISE LINEAR REGRESSION. All of the rise time and $50\%_{max\_amp}$ values were used as predictors for the P-centres. The best predictor of the P-centres was the channel 5 $50\%_{max\_amp}$ values ($s=12.2$, $r^2=53.05$).

The fit of this regression was tested using BEST SUB-SETS REGRESSION. All the channel output $50\%_{max\_amp}$ values were used as predictors in the analysis. The $50\%_{max\_amp}$ values for channel 5 were the best predictors, overall the possible sub-sets of channel outputs, giving the fit nearest to zero ($C\text{-}p=2.1$).

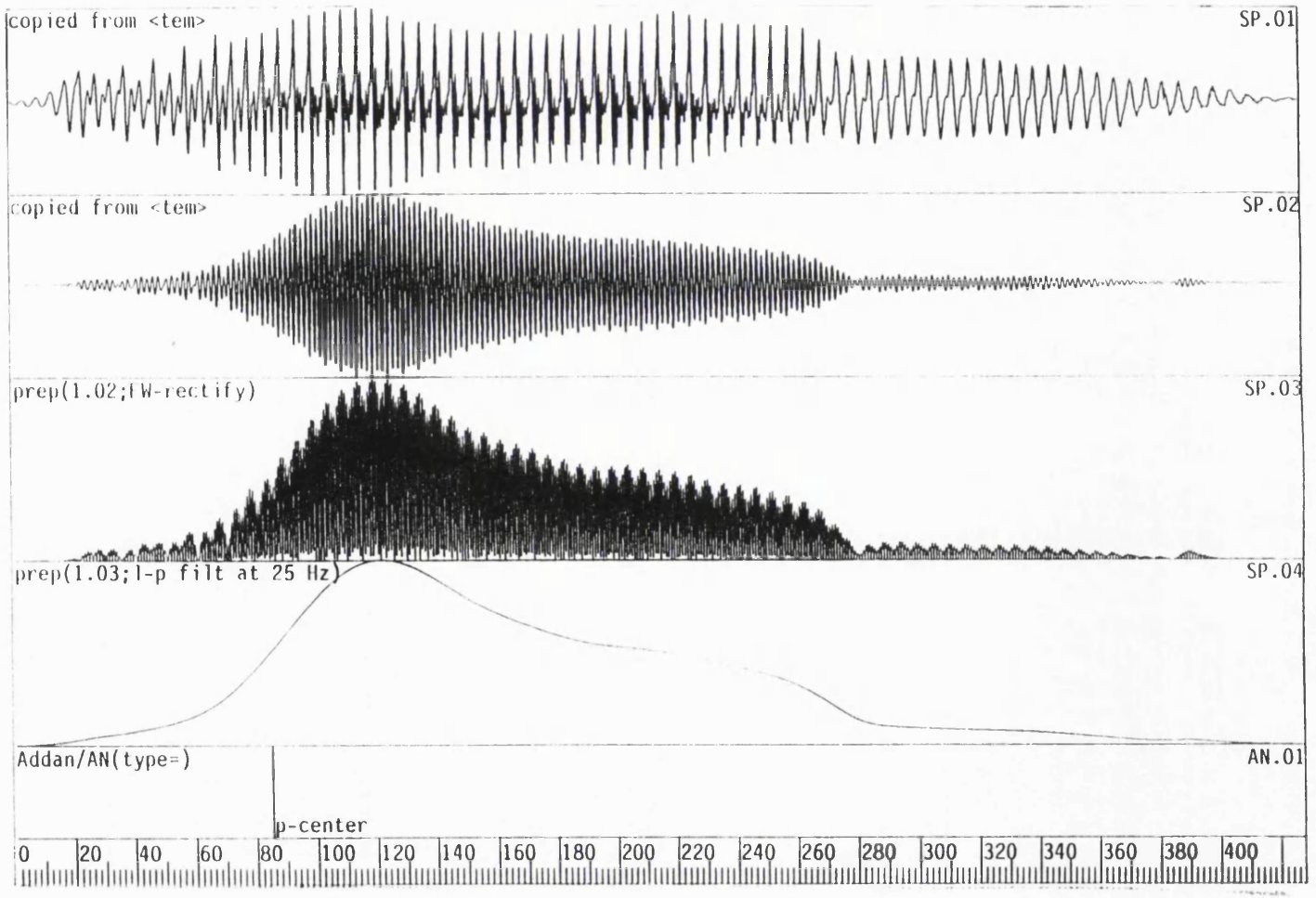


Figure 12.6 Output of FAIM P-centre model; original signal, channel 5 GammaTone filterbank output; full rectification; smoothing (Butterworth filter at 25Hz); P-centre annotation at 50%$_{max\text{-}amp}$

A linear regression (fitting the model $y = c + \alpha x$) of P-centre against the predictor $50\%_{max\_amp}$ of channel 5 is significant ($F_{1, 14}=15.82$, $p<0.05$). The regression equation is:

$$\text{P-centre} = -11.2 - 0.407(50\%_{max\_amp})$$

The $50\%_{max\_amp}$ value is therefore a significant predictor of P-centre location. In the frequency dependent onset attributes described earlier in this chapter it was described how the increases in amplitude, and the nature of these increases were to be modelled. Thus both changes in the amplitude of a frequency region **and** the quality of the increases in amplitude in this region are hypothesized to be important in the location of P-centre. It is therefore necessary to establish that the rise time of this channel output varies with the P-centre.

This was tested by correlating the P-centres against the rise time values and the $50\%_{max\_amp}$ values for the channel 5 output.

The P-centres correlated with the channel 5 $50\%_{max\_amp}$ values significantly ($r=-0.728$, $df=15$, $p<0.01$).

The P-centres also correlated significantly with the channel 5 rise time values to a significant amount ($r=-0.538$, $df=15$, $p<0.05$).

Thus both the 'location' of this increase in amplitude (the $50\%_{max\_amp}$ values), and the quality of this increase (the rise time values) correlate with P-centre location.

When the two predictors, rise time and $50\%_{max\_amp}$ values for channel 5, were correlated together, the correlation coefficient just failed to reach significance ($r=0.451$, $df=15$, $p>0.05$). This indicates, as outlined previously, that the two parameters do not necessarily correlate. In Chapter 13, the conclusions

chapter, the possibility of using one parameter which captures both rise time and $50\%_{max\_amp}$ is discussed.

12.6 A Model of P-centre location

Figure 12.6 shows the output of the stages of the FAIM model, filtering, rectifying, smoothing and P-centre annotating, for the "two" speech item from speaker SR. The integration of location and slope of the amplitude increase within this frequency bandwidth can be seen. The centre frequency of this channel is 578Hz. It is increases in amplitude within this channel frequency bandwidth which are leading to the perception of P-centres. Note that this is a similar frequency region to that used by Marcus (1981) in his model of P-centre location to define vowel onset (vowel onset was defined as the peak increment in amplitude between 500-1500Hz). The role of vowel onset was central to Marcus's model, which partitioned the syllable into two sections according to vowel onset, and related P-centres to the durations of the different portions. The proposed model of P-centres in this thesis can be regarded as shifting the emphasis of Marcus's model away from the durations of segments within a syllable, towards the development of perceptual structure within a syllable, which arise due to rapid increases in specific frequency channels. The model will be referred to as the Frequency dependent Amplitude Increase Model of P-centre location, or FAIM.

12.6 FAIM vs. Marcus's Model

Although there are conceptual similarities between FAIM and Marcus's P-centre model, there are also many differences. As described in Chapter One, Marcus's model requires that the signal passes an arbitrary threshold to determine the onset and offset leading to the possibility that a sequence of speech signals all about this threshold could contain only one P-centre; FAIM is not as dependent

upon the baseline amplitude level. Marcus's model thus involves some prior segmentation of the signal into onset/offset chunks; FAIM requires no such processing. Marcus's parsing of the syllable structure into pre- and post- vocalic segments is relevant for speech signals but not for non speech signals; the approach basic to FAIM could be applied to any acoustic signal.

To test whether this model of P-centre location - the FAIM description - can make reasonable predictions, an implementation of the model was used to generate predictions for the P-centres for all the stimuli which have been used in this thesis. The code for this implementation is given in the APPENDIX.

To test whether the FAIM predictions are comparable to those of Marcus's model, the implementation of his model described in Experiment 3 was used to make predictions for the P-centres of all the stimuli tested in this thesis. If the FAIM model performs more accurately as Marcus's, or as well, it suggests that P-centres can be modelled locally, without any explicit duration information; P-centres would thus be less a function of the entire syllable, than of the perceptual structure of the syllable and acoustic events within it.
Table 12.1 shows the P-centres, FAIM predictions, and Marcus model predictions for all the thesis stimuli.

To test statistically which model predictions best fitted the experimental P-centres, a BEST SUBSETS REGRESSION was performed. This returns the predictor which gives the best fit - the C-p value. The nearer this is to zero, the better the fit. $C_p \quad \frac{SSE_p}{MSE_m} - (n-2_p)$

The best predictor was the FAIM predictions (s=15.94, r^2=85.8). The C-p value was 7.9.

The Marcus model predictions (s=25.26, r^2=64.3) gave a fit of C-p=77.2. This is considerably worse than the FAIM fit.

The FAIM predictions give the best fit to the experimental P-centre data; to test the significance of the relationship the P-centres were regressed separately against the FAIM predictions and the Marcus Model predictions using linear regression.

1) FAIM as predictor
The predictor FAIM was a significant predictor of P-centre in linear regression ($F_{1,40}$=241.36, $p<0.05$).

The regression equation is:

P-centre = 1.22 + 0.615 FAIM predictions

2) Marcus model as predictor
The Predictor Marcus model predictions was significant ($F_{1,40}$=72.02, $p<0.05$).

The regression equation is:

P-centre = 15.8 + 0.400 Marcus Model Predictions

The statistical test results therefore show that the FAIM predictions account for more of the variance in P-centres, and give a better fit, than the Marcus model predictions. The Figures 12.7 and 12.8 show the P-centres plotted against the FAIM predictions and the Marcus model predictions respectively.

speech stimuli	P-center(ms)	FAIM predictions (ms)	Marcus model (ms)
ramped "ae"	-3.7	-11.5	-53.3
	-15.2	-43.3	-59.3
	-24.0	-52.2	-77.3
ramped "wa"	-43.2	-75.0	-258.3
	-65.2	-84.3	-258.3
	-71.5	-84.7	-258.3
duration of frication	-155.2	-228.2	-328.0
	-111.5	-168.2	-289.0
	-62.2	-108.3	-25.0
ramped "sha"	-148.5	-228.2	-328.0
	-152.5	-228.2	-328.0
	-152.0	-228.2	-328.0
speech from SKS	-55.2	-107.8	-224.3
	-45.0	-113.9	-212.7
speech from SHS	-26.6	-63.1	-222.3
	-31.6	-86.4	-175.3
speech from SR	-43.0	-85.4	-201.2
	-47.0	-104.7	-194.7
speech from SM	-3.3	-37.6	-132.2
	-29.6	-25.5	-108.3
speech from PH	-70.6	-68.1	-179.5
	-30.6	-0.2	-140.6
speech from LCE	-11.6	-9.7	-146.7
	-17.8	-77.7	-108.6
speech from WC	-49.0	-77.2	-129.0
	-52.0	-104.5	-143.1
speech from DG	-31.6	-67.8	-97.3
	-40.3	-5.4	-92.6
peak clipping expt; normal speech	-67.7	-107.4	-176.4
	-51.5	-116.9	-146.8
	-78.9	-126.1	-90.6
	-32.2	-75.0	-121.8
peak clipping expt; $speech_{ipc}$	-46.3	-101.6	-142.0
	-18.9	-24.6	-130.4
	-37.4	-12.7	-136.4
	-9.4	-19.8	-96.5
ramped synthetic vowel onset stimuli	+0.4	-0.7	-50.0
	-4.2	-18.6	-58.0
	-10.2	-31.9	-76.0
	-16.0	-44.8	-76.0
$eight/eight_{burst}$ stimuli	-19.0	-35.3	-123.3
	-12.0	-35.3	-123.3

Table 12.1 P-centers, FAIM and Marcus model predictions for all thesis stimuli.

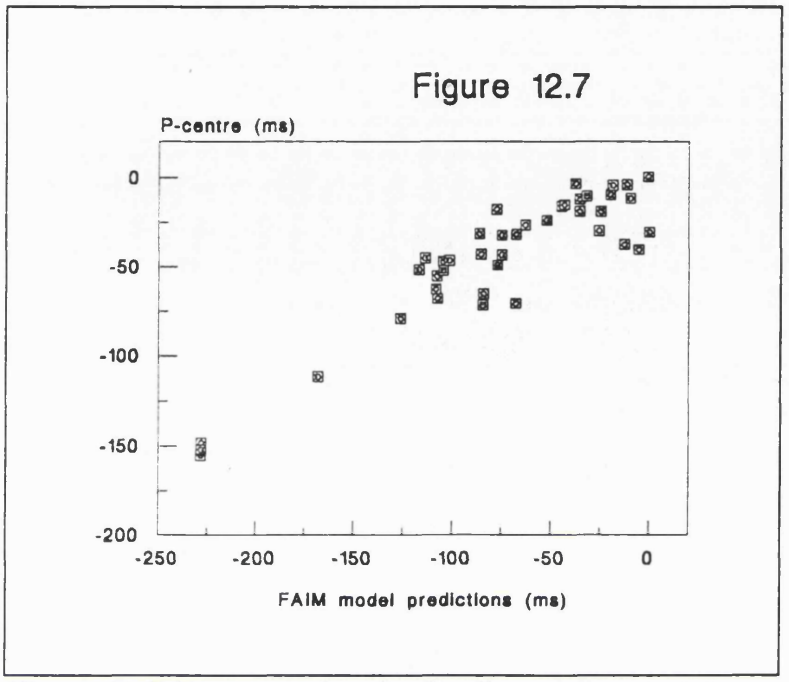


Figure 12.7 P-centres for all stimuli examined in this thesis plotted against FAIM predictions

Figure 12.7 shows that, as the statistical test results suggest, the FAIM predictions are a good fit to the observed P-centre data. The relationship is linear.

The Figure 12.8 shows that the Marcus model predictions are not such a good fit for the P-centres. The predictions of this model appear to fall on a curve; this is the reason why the Marcus model predictions are a less strong P-centre predictor in a linear regression.

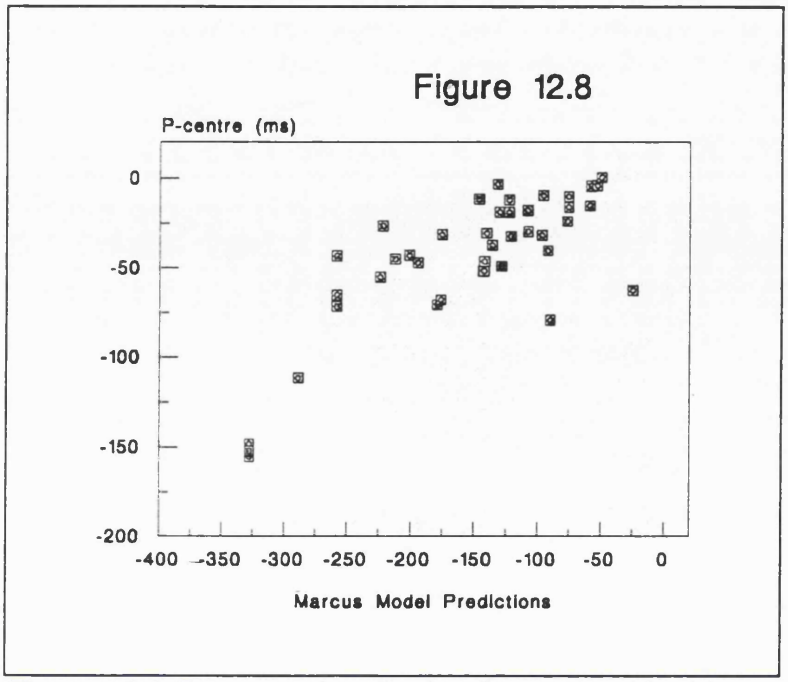


Figure 12.8 P-centres for all stimuli examined in this thesis plotted against the Marcus model predictions

12.6 Discussion

The implications of this comparison are clear. The two models are similar in their predictions (and correlate together significantly, $r=0.764$, $df=41$, $p<0.05$). The FAIM model is the better model in terms of accounting for the observed P-centre variance, and goodness of fit. The relationship between P-centres and Marcus model predictions is non linear.

This similarity of the two model's output is expected since they both are based around a similar parameter - the increase in mid-band energy. As explained earlier however, the FAIM model has a different perspective to Marcus's model, relating P-centers to the perceptual consequences of this increase in energy.

The FAIM model could be considered to describe why Marcus's model does make reasonable predictions about P-centre location. Marcus's model was originally developed as the best fit to the data, rather than a theoretically driven model.

The FAIM model is based upon the theoretical principle of onset events leading to perceptual events from the musical perception literature (Gordon 1987, Vos and Rasch 1981), and experimental results. The physical development of the model was performed using natural speech from a variety of speakers, to test the hypothesis that the theoretical principle could be applied to free speech sounds.

The FAIM model is also driven by the hypothesis that it should be possible to model P-centres in a local manner - as opposed to a global syllable structure approach. This approach would account for the perception of P-centres in entirely novel sounds in real time (where previous models require the entire signal to be processed in order to perceive the centre). If the increases in energy relate to attributes of the onset of vowel articulation, then P-centres in this model would map into articulatory actions, as Fowler first suggested (Fowler 1979).

12.7 FAIM and Howell's criteria for a determinant of P-centre location

As was described in Chapter 1, Howell (1984) defined three criteria for a determinant of P-center location. He did this with reference to a determinant in a global model of P-centre location. The FAIM description of P-centres is a local model, and attempts to identify an acoustic event to which P-centres

correspond directly. In this section Howell's criteria will be applied to the FAIM approach, to test whether the proposed model fulfils all three.

1) *"it should vary in alignment across stimuli in the same way as the perceptual judgements do" (Howell 1984, p. 429)*

The aim of this Chapter was to develop a model of P-centres which was based upon stimuli whose P-centres had been determined in a dynamic rhythm setting task. When the set of sixteen speech items from eight different speakers was analyzed, the location and rise time of the amplitude increase in the channel centre frequency 578Hz were significant predictors of P-centre location. This indicates that this frequency specific amplitude increase is varying across the natural stimuli as the P-centres - reflecting subjects' perceptual judgements - vary across the stimuli.

2) *"it should vary in alignment relative to stimulus onset in the same way that perceptual alignments vary when the acoustic properties of test stimuli are altered" (ibid)*

As described above, this Chapter (and the thesis) was designed to develop a model of P-centre location. The predictions made by FAIM have not been tested experimentally; instead the model has been applied to stimuli to determine whether the model predicts the experimentally measured P-centres. This section thus considers how the predictions of FAIM can be applied to the results of experiments where the speech stimuli were systematically altered. This is thus a *post hoc* interpretation of the results.

In Experiment 6, when the onsets of three types of speech sounds were ramped, the FAIM parameters were altered in two of the three speech types - the ramped semi-vowel ("wa") and the ramped vowel ("ae"). The FAIM

parameters were unchanged in the third ramped speech type - a fricative ("sha"). Corresponding to this, the P-centers of the "wa" and "ae" stimuli shifted away from the onset of the syllables, with the increased amount of ramping; the P-centers of the ramped "sha" stimuli did not vary significantly with the increased amount of ramping.

In Experiment 4, the infinite peak clipping of natural speech sounds resulted a moving of the FAIM parameters nearer to the onset of the signal. This shift is replicated in the P-centre change caused by infinite peak clipping; for every speech item the P-centre was moved towards the onset of the syllable.

This indicates that the FAIM parameters vary in relation to the onset of a syllable in correspondence with the subjects' settings when speech items are acoustically manipulated.

3) *"the factor should account for....(P-centres)....in both production and perception" (ibid)*

Again this criterion was not addressed in this thesis. In this section the hypothesis that FAIM could be applied to speech production will be considered.

In this thesis only the perception of P-centres and factors which affect them were experimentally investigated. Some production data was collected; this was principally used to compare to the perception results (eg. Experiment 2b). No articulatory data (eg. EMG measurements) was collected. Therefore no empirical evidence can be given which would support FAIM as accounting for P-centres in production.

It could be hypothesized however that P-centres, when modelled as local acoustic events, reflect articulatory actions in the production of speech. As

suggested earlier the FAIM account of P-centres may reflect the manner in which the onset of a vowel is articulated. A local model of P-centre location, based upon vowel articulation, was put forward by Fowler (1979); however her approach precluded the possibility of defining acoustically where P-centres are in the signal.

The P-centre differences across speakers with different British English accents (Experiment Three), which are predicted by FAIM, appear to be based upon varying articulatory patterns across speakers. Listening to the speech from the different speakers, it is apparent that some produce the word "one" with gentle onsets (eg. speakers PH and SKS, both from the North of England); others produce "one" with an abrupt onset (eg. speaker LCE, from Hampshire).

Thus for Howell's final criteria; the FAIM approach does account for P-centres in perception and **may** provide a framework for considering the role of discrete articulatory gestures in P-centre production.

12.8 FAIM and Pompino-Marschall's model of P-centre location

The FAIM and Pompino-Marschall models (described in Chapter One) are similar in the frequency dependent, psychoacoustic attributes of their approaches; they differ more in the underlying concepts of what P-centres are. In Pompino-Marschall's model they are the result of other perceptual attributes connected with the syllable and can vary across subjects. In the FAIM model they are punctate perceptual events which can be mapped directly onto the onset of amplitude increases in the frequency region around channel five.

12.9 Summary

An approach for modelling P-centres based upon the experimental evidence was described; a modelling protocol was outlined and suitable analysis tools, parameters and stimuli chosen. Using the APU GammaTone cochlear filterbank model to provide an auditory image of the speech sounds (4.0 ERB channel spacing), the $50\%_{max\_amp}$ values of the channel centre frequency 578Hz were found to be significant predictors of P-centre location. To test whether this parameter could be used to model P-centres generally, an implementation of this Frequency dependent Amplitude Increase Model (FAIM) was used to generate P-centres predictions for all the stimuli used in this thesis. These were compared to the experimentally determined P-centres. The FAIM implementation predicted a significant amount of the observed variance (r^2 = 85.8%). This was compared to the predictions of the Marcus model for the same stimuli; this made reasonable predictions but accounted for less variance. The similarities between the FAIM and Marcus models were discussed. Similarities and differences between Pompino-Marschall's model and FAIM were discussed.

Chapter Thirteen
Conclusions

Abstract
The experimental work carried out for this thesis is reviewed, and the FAIM model situated in the context of this work. The implications of the experimental work described in this thesis for other models of P-centre location are discussed. The relevance of P-centres and this study to other areas of research is considered - for example the impact of this on auditory scene analysis, and a consideration of some issues not addressed in the thesis.

13.1 Introduction

The aims of this thesis as stated in Chapter 3 were to evaluate current P-centre models, test intensity and time parameters in terms of their effects on P-centre location and develop a model which quantifies P-centres as in terms of auditory events which are non speech specific. The framework of the model was derived from current models of musical beat perception, in an attempt to reconcile differences between P-centre models in the speech and non speech literature.

This approach involved experiments which tested some assumptions concerning the P-centre hypothesis, which produced data against which various P-centre model predictions could be compared. Further experiments were performed to manipulate possible modelling variables.

13.2 Experimental work

13.2.1 EXPERIMENT ONE was designed to ensure that subjects could perform the dynamic rhythm setting task by which P-centres are measured. A group of twenty naive subjects set the experimental reference sound to a perceptually even rhythm. The variables manipulated were tempo and whether or not the

subjects were allowed to make movements to help their settings. The results would therefore indicate whether subjects could do the task, whether a faster or slower tempo made a difference, and whether 'tapping' along aided accuracy. The results were analyzed in terms of how accurate the subjects were in terms of the deviations of their settings from physical isochrony. The slower tempo increased subjects' accuracy, and tapping improved subjects' accuracy in the faster condition. One subject's results were rejected due to a failure to acheive any isochronous settings. "Subject" was otherwise not a significant predictor of settings, indicating that subjects do not differ in their ability to make the settings. In the rest of the thesis experiments a slower tempo was used, and subjects were allowed to make any actions they felt aided their settings.

13.2.2 EXPERIMENT TWO was designed to test the hypothesis that the pattern of timing found when speakers produce isochronous speech is analogous to the pattern of intervals aligned by subjects when setting stimuli to perceptual isochrony. This therefore tests the hypothesis that the P-centre phenomena (of perceptual isochrony not being due to physical isochrony) is found in both production and perception. In this experiment seven speakers produced isochronous speech sequences *at the tempo which was most natural to them*; the patterns of timing for each speaker were collected. These were compared to the intervals subjects set when instructed to align items from each speakers' speech to an isochronous rhythm. Naturally timed isochronous speech sequences were thus compared to the intervals set in a dynamic rhythm setting task. The ratio of intervals in production was a significant predictor of the ratio of intervals set in the perception experiment. This is despite the tempo of the perception experiment being controlled, and the tempo of the production section varying according to speaker. The precise ratios in perception and production varied across speakers, indicating that the different speakers' speech varies in P-centre location.

13.2.3 EXPERIMENT THREE tested the hypothesis that the variation in interval ratios found across the different speakers reflected varying P-centre location across the speech. P-centres were determined in full rhythm setting tasks for eight speech items from four speakers. The P-centres were found to vary in line with the original pattern of production intervals. Implementations of two P-centre models and one Perceived Onset Time model (Howell 1988a, Marcus 1981, Vos and Rasch 1981) were used to make P-centre predictions; both the Howell and the Marcus models accounted significantly for the observed P-centre variation. Additional P-centres were established for the speech of the four remaining speakers in reduced rhythm setting tasks, and these were found to relate to the original production interval patterns. P-centres therefore do underlie the timing of isochronous speech across speakers.

13.2.4 EXPERIMENT FOUR was an attempt to replicate Tuller and Fowler's 1981 finding that rendering the amplitude envelope of a syllable invariant by infinite peak clipping does not affect the P-centre location. Speech was infinitely peak clipped digitally, and reproduced at the subject's ear via equipment which did not distort the phase or amplitude of the signal. Subjects performed alignment tasks, setting stimuli, either normal or infinitely peak clipped to a rhythm against the reference sound. The deviations from absolute isochrony for the infinitely peak clipped speech were always smaller than those for the normal speech, indicating that in every case the P-centres had shifted towards the onset of the syllable. The model implementations used in **EXPERIMENT THREE** were used to generate predictions for the stimuli; both the Vos and Rasch Perceived onset model and the Marcus P-centre model accounted significantly for the observed variation in deviation from isochrony. The Vos and Rasch model made good predictions for the speech stimuli, but not the $speech_{ipc}$ stimuli, since it is noise sensitive. The Marcus model was a significant predictor of the settings, despite this model not incorporating any explicit amplitude information. This was found to be due to the definition of vowel onset

used in Marcus's model (peak increment in mid-band spectral energy), which was affected by the infinite peak clipping.

13.2.5 EXPERIMENT FIVE was designed to test Howell's (1988a) model of P-centre location. This model predicts that increasing the amplitude at the offset of a syllable will shift the P-centre of that syllable *towards* the offset. The experiment was a replication of Marcus's (1981) finding that increasing the amplitude of a syllable final "t" burst in the word "eight" did not affect the P-centre of the syllable. The results indicated that, as Marcus found, the manipulation did not affect the P-centre. The P-centre of the "$eight_{burst}$" stimulus appeared to shift slightly towards the onset of the signal; this result seemed to be due to difficulties in setting this stimulus to a rhythm as the amplified "t" burst tended to syllabify. In addition, Marcus's finding was extended to the production of speech; speakers were instructed to alternately increase the "t" bursts when uttering "eight" rhythmically. Howell's model predicts that since the production of speech is timed by speakers according to the energy distribution of the syllables, this instruction should affect their timings (Howell 1984, 1988a). The results showed that the increased alternate "t" bursts did not affect the subject' timing as this model predicted. The syllabic centre of gravity model of P-centre location was thus not a predictor of timing in production and perception.

13.2.6 EXPERIMENT SIX manipulated the rise time of a signal to examine the effect on the P-centre. This is the manipulation that Vos and Rasch (1981) found affected the Perceptual Onset Time of musical note, and that Howell (1984) found affected the alignment of both speech and non-speech sounds in a quasi-rhythm setting task. This experiment involved the ramping of the onsets of different types of speech sound, in the light of previous experiments suggesting that onset events are more important than offset events in determining P-centre location. A CV syllable (where C = "sh") a CV syllable

(where C = "w") and a vowel were ramped to different amounts, and the P-centres determined. The effect of ramping the onsets varied according to the identity of the speech items - ramping the vowel and the semi-vowel CV syllable ("wa") affected the P-centres, to comparable amounts; ramping the fricative CV syllable ("sha") had no significant effect. This result indicated that ramping the onset of a syllable affects the P-centre - if the segment ramped has certain qualities, which could be expressed in terms of the perceptual loudness, or sonority of the speech sound. Ramping more sonorous sounds affects the P-centre of the whole syllable, ramping less sonorous sounds (eg."sh") does not affect the P-centre. Both the location in the syllable, and the perceptual quality of the speech segment ramped affects whether the P-centre of that syllable is shifted.

13.2.7 EXPERIMENT SEVEN tested the hypothesis that onset events are more important than offset events in determining P-centre location, in the light of the previous result showing that not all speech sounds are perceptually equal in terms of the effects of their amplitude manipulation. That is, the results of **EXPERIMENT FIVE** showed that increasing a syllable final "t" burst amplitude did not shift the syllable P-centre, but this might be a function of "t" being of low sonority. In this experiment therefore the onset and offset of a speech sound were ramped to create two families of stimuli. The stimulus was a synthetic vowel of constant pitch to ensure there was no difference in the identity of the segments ramped. The results showed that, as in **EXPERIMENT SIX** ramping the onset of the vowel-like sound shifted the P-centre away from the onset of the sound. Ramping the offset had no significant effect on the P-centres. Therefore the conclusion was that onset events are more important in determining P-centre location than offset events.

13.2.8 EXPERIMENT EIGHT tested the hypothesis that vowel duration affects P-centre location. The previous experimental evidence in support of this has

been mixed, with differences both between experiments in the size of the effect reported, and between subjects in the within experiments. If, as is hypothesized in this thesis, it is possible to model P-centres locally (as are musical notes) then P-centres need not be affected by the vowel duration. The experimental stimulus was a synthetic vowel with constant pitch, edited to five different durations to avoid the differential effects of shortening natural vowel sounds. The results indicated no significant effect of vowel duration on P-centre location. The previous variation in vowel duration effect was considered in terms of Lame's (1990) hypothesis that dynamic rhythm setting tasks are affected by segment duration not as a function of P-centre shift, but due to temporal illusions. On the basis of this lack of significance, it was decided to attempt to model P-centres without incorporating vowel duration information.

Table 13.1 indicates the different P-centre models, and how they predict the experimental findings described in this thesis. A check mark indicates that the model predicted the result, through implementation or model structure; a cross indicates that the model did not predict the result; a question mark indicates that the model had no concrete predictions to make on this experimental manipulation.

	Model predictions			
Experimental result	Marcus model	Syllabic centre of gravity	Articulatory	Vos and Rasch
differences between speakers	✓	✓	?	×
infinite peak clipping shifts P-centre	✓	×	×	✓ but noise sensitive
altering $Eight_{burst}$ does not affect P-centre	✓	×	?	✓
Ramping some onsets shifts P-centres	✓	✓ not frequency effects	×	✓
ramping vowel onset shifts P-centre - ramping offset does not	×	× not offset effects	?	✓
vowel duration effect slight	×	×	×	× not addressed in model

Table 13.1 The different P-centre models and their predictions of the P-centre experiments described in this thesis - a check mark indicates that the model did predict the result, a cross that it did not, and a question mark indicates that the model did not address that experimental manipulation.

13.3 Modelling P-Centres - the FAIM model of P-centre location

A model of P-centre location was developed based upon all the experimental results. It was hypothesized that since P-centers were affected by the rise times of certain vowel sounds, that P-centres might be due to rapid increases in amplitude of frequency regions corresponding to very sonorous speech sounds, ie. vowels. This corresponds to the success of Marcus's (1981) model, which is due to his particular definition of vowel onset (peak increment in mid-band spectral energy). A model was developed which did not look for vowel onset *per se*, but frequency region amplitude increases, which may or may not correspond to vowel onset. This is thus satisfying the stated aim of modelling P-centres in a parsimonious, non speech specific, punctate manner. The model was developed using speech items from eight different speakers, and when tested against all the stimuli in the thesis, was a better predictor than Marcus's model. The FAIM model can be regarded as supplying a theoretical explanation for why Marcus's original model (which was developed on a data driven basis) generally accounts for data well; it can also be regarded as successfully describing P-centres in the same conceptual framework as the models of musical beat location. P-centres in FAIM are the auditory consequences of increases in perceptually salient frequencies.

13.4 FAIM and P-centres

The relationship between the acoustic structure of speech sounds and the perception of rhythmic centres is quantified in this thesis as resulting from increases in the amplitude of certain frequencies. These frequencies correspond to those of the vocalic onset portion of a syllable. P-centres in this model thus arise from perceptual events associated with the vowel onset, and

which are affected by the qualities of the vowel onset. This perspective of P-centres is very different to several previous conceptions (Cooper et al 1986 1988, Howell 1984 1988a&b, Pompino-Marschall 1991), in that it does not require the entire signal to be processed before the P-centre is detected. It also differs from Fowler's 1979 definition of P-centres as the onset of vowel articulation, since she defined this as not being expressed in the acoustic waveform (see Chapter One for further details). P-centres are conceived of as being due to identifiable acoustic / perceptual events within the signal. P-centres in this framework represent perceptual non-linguistic structures in continuous and metrical speech sequences.

The FAIM model, as defined in Chapter 12, is motivated by models from music perception literature which model beat location in terms of the onset amplitude increases; the same parameters which were identified as perceptually important in Chapter 2. The rise time characteristics of FAIM place it in the same framework as these non speech models; the frequency specific attributes of FAIM are the basis of its similarity to Marcus's original model. This thesis has therefore acheived one of its stated aims, to reconcile the difference between non speech local beat models, and speech based global P-centre models. In addition, and rather satisfyingly, the model derived from the thesis has also indicated that Marcus's original model was a good one and derived perceptual reasons why it worked, despite the non theoretical development.

13.5 Implications for other models of P-centre location

13.5.1 Marcus's model The implementation of the Marcus model described in **EXPERIMENT THREE** provided good predictions of the observed P-centre variance. This was found in experiments where the manipulations apparently affected stimuli parameters which did not feature in his model, for example the effects of infinite peak clipping in **EXPERIMENT FOUR**. As was described in

Chapter 12, the Marcus model was a mathematical fit to the observed data, with no theoretical basis.

The FAIM model has most in common with Marcus's original P-centre model, as was described in Chapter Twelve, due to the Marcus model definition of vowel onset and the importance of this parameter within the model being similar to the sole parameter in FAIM. As was described in Chapter Twelve, FAIM can be regarded as altering the emphasis of Marcus's model away from the duration of syllable segments and onto qualities of the vowel onset. FAIM thus provides a perceptual basis for the function of Marcus's model.

13.5.2 Musical Beat Models - Gordon (1987), Vos and Rasch (1981) The work in this thesis, as mentioned in the previous section, aimed to resolve the discrepancies between the theories of beat location in the speech and non speech literature. This aim has been acheived; the same parameter that Gordon and Vos and Rasch identified as determining perceptual beat location (rise time) has been successfully applied to perceptual centres in speech. The implementation of the Vos and Rasch model was however noise sensitive, which often affected its predictions about the P-centres of speech stimuli (eg. **EXPERIMENT FOUR**). In order for a musical model approach to work therefore speech signals must first be filtered; it is the high sonority, mid band frequency components rise times that determine P-centre location in speech sounds.

13.5.3 Centre of Gravity hypothesis The work in this thesis did not support the centre of gravity hypothesis of P-centre location (Howell 1988a). This model of P-centres is equally influenced by the amplitude variance over the whole signal duration; the experimental work in this thesis showed that onset characteristics determine P-centres. The centre of gravity model weights all frequency components equally; the work in this thesis indicates that not all frequency contents are as important in determining P-centres. The model might

be altered such that the duration/intensity relationship was differentially weighted, and the signal filtered to reduce the frequency equality. The centre of gravity model of P-centre location cannot as it stands account for the observed results. This has been found previously (Seton 1989).

13.5.4 Articulatory theories The work in this thesis has only directly addressed the articulatory hypothesis in one experiment. The experimental work has been more directed to developing a general model than testing the articulatory approach, although the conceptual basis for the FAIM approach is antithetical to articulatory models. Specifically, the results of **EXPERIMENT FOUR** showed that the amplitude / time characteristics of a signal do alter its perceptual centre, despite the intact linguistic information (the speech being still identifiable). This does not support the articulatory hypothesis that P-centres are directly perceived from the speakers utterance gestures, and cannot be influenced by the physical signal characteristics. The perception of gestures may well affect the P-centre location, but these gestures are represented in the physical signal and can be affected by manipulation of that signal.

The overall construction of FAIM supports this non direct perception concept. The fact that P-centres are being modelled in a non speech specific manner, which is based upon the physical signal, is unacceptable to the direct action perception theorists; it is unlikely that this dichotomy will be resolved over P-centres.

13.5.5 P-centres and Lame's hypothesis As part of his thesis on internal clocks, Lame (1990) suggested that a principle assumption of P-centres - that they are context independent (Morton et al 1976) - is incorrect. P-centres, he argues, like other rhythmic phenomena are influenced by the interval perceptions, and thus by the content of the other signals in the rhythmic sequence. This was his explanation for the effect of vowel duration on P-centre

location (noted by Cooper et al 1986 1988, Fox and Lehiste 1987, Marcus 1981). Lame's hypothesis was not tested in this thesis, although the results of **EXPERIMENT EIGHT**, where a the duration of a synthetic vowel was altered and caused no significant P-centre change, suggested that he could be correct (because it seems, the simpler the duration varying stimulus, the less of an effect on P-centre). Evidence to support Lame's assertion that P-centre are not context independent comes from Seton (1989); he found a difference on the effect of the intensity of stimuli on their P-centres in mixed or blocked loudness presentations. This again indicates that the other stimuli in an dynamic rhythm setting task - the context - affect the results of the experiment, and that a subject's performance in such a task may be based upon the P-centres of the stimuli, but also affected by other experimental aspects.

13.6 Why model P-centres?

P-centres represent the perceptual moment of occurrence of acoustic signals, and are thus of relevance to several areas of psychological research. The FAIM model could be applied to various types of research, either as a psychologically plausible preprocessor of incoming acoustic stimuli (eg. for a rhythmic parser), or as a possible explanation for certain research findings.

13.6.1 The study of synchronous action mentioned in Chapter Five has consistently shown that perceptual synchronicity is not based on physical synchronicity (Auxiette and Gerard 1992, Prinz 1992, Semjen, Schulze and Vorberg 1992), suggesting that the concept of perceptual moment of occurrence can be applied to the perception of produced actions, as well as perceived and produced acoustic signals. The study of P-centres in this framework is thus the study of the correlation between action and perception across sensory domains (eg. auditory and kinaesthetic).

The pattern of responses which are commonly found in synchronization tasks indicate that subjects achieve perceptual simultaneity (when tapping along to a rhythmic stimulus) by tapping **before** the external acoustic stimulus. The study of P-centres indicates that this difference between perceptual and physical synchronicity and isochrony is dependent upon the acoustic qualities of the external stimulus in perception, and the inertial structure of the system which produces the action in production. P-centres may represent a different approach to the synchronization of action with sound, of other synchronization tasks across different action domains, by focusing attention upon the attributes of the produced action / perceived signal. It could be hypothesized that if P-centres exist as perceptual structures in sound, then perhaps other motor actions (eg rowing, clapping) have a perceptual moment of occurrence, or perceptual centre.

13.6.2 P-centres - the perception of an acoustic 'beat' - can be seen as the primary stage of rhythmic parsing of acoustic input, and are thus of relevance to the study of rhythm perception (see Lee 1991 for a review of this research). A rhythm can only be perceived if 'beats' arise perceptually in the acoustic signal, and the evidence suggests that P-centres are analogous to perceptual beats in this context. Models of human rhythmic parsing could thus utilize P-centres as a perceptually based first pass processing of the rhythm. For example the model of rhythm perception of Lerdahl and Jackendoff (1983) involves a preference rule (MPR 3) about the coincidence of note-onsets with beats (beats in this sense referring to the metrical structure of the constructed rhythm). There is evidence that such a rule is not *per se* sufficient for determining the metrical structure of a sequence (Lee 1991). P-centres however represent a perceptually appropriate processing of the note-onset stage of any synchronization based rule about rhythm perception.

13.6.3 Languages have been construed as either stress timed (eg. English), syllable timed (eg. Spanish) or mora timed (Japanese). In stress timed languages, the intervals between stressed syllables (stress-feet) are held to be constant. The evidence for such a global rhythmic structure in spoken English has generally shown that there is a tendency towards isochrony (Hill, Jassem and Witten 1978). This has led to the hypothesis that evenly timed stress-feet rest more on our perceptions of speech than any strict rhythm in the input (Lehiste 1977, Miller 1984, Scott, Isard and de Boysson Bardies 1985). Thus researchers have concluded that overall there is little evidence for stress based isochrony in speech, and that speech is timed on a word by word basis. P-centres have been hypothesized to represent a 'bottom-up' approach to timing in speech, one which is constrained by the rhythmic properties of the speech signal rather than an overall rhythmic structure (Seton 1989).

Finally, P-centres represent the perceptual phenomena that acoustic events are not undifferentiated over time, but have a perceptual focus, or event which provides the signal with a perceptual structure. In speech P-centres represent perceptual structures within syllables associated with vocalic onset. P-centres therefore are a method for investigating the structure of our acoustic environment, in relation to the external stimulation. The study of P-centres has a place therefore in the framework of auditory scene analysis (Bregman and Campbell 1971), providing a process for event detection in parallel to stream segregation.

13.7 Directions for research

Inevitably the research in this thesis raised almost as many questions as it answered. In this section topics not fully addressed are outlined.

13.7.1 The relationship between P-centres, frequency regions and amplitude change could be elaborated with further research. This would refine the definition of the FAIM, perhaps using more speakers, and more speech sounds. Ideally P-centres would be modelled using a variety of speech items. The rise time parameter is currently defined with two variables, the rate of the rise in amplitude, and the location within the syllable of the increase in amplitude. These two variables could be unified into a single parameter, which expressed both. A hyperbolic curve function of the amplitude rise time would contain both types of information, and might make a more suitable parameter to base the FAIM implementation upon. This would also provide a frame work for considering explicitly the rate of change of amplitude that gives rise to the perception of an acoustic event, the assumption that this amplitude change affects perception being implicit in most acoustic models.

The relationship between frequency region and perceptual salience could be extended; signals which do not contain the frequencies from channel 5 would still be hypothesized to have a 'moment of occurrence'. The nature of the interaction between spectral contents and P-centre location could thus be amplified and incorporated into the model as frequency weighting.

13.7.2 Further research could also address the status of P-centres in perceptual processing. How 'early' a process is the determination of P-centres in terms of the parsing of the acoustic input. This is relevant experimentally, since there is a suggestion that the P-centres of stimuli are affected by the degree to which the stimuli in a dynamic rhythm setting task can be streamed together. This was noted by Seton (1989) who found that subjects had great difficulty in making rhythm settings if the stimuli were separated in pitch and thus hard to stream together. In addition he found a slight shift in the P-centres under such conditions, though whether this was due to a genuine P-centre change or experimental difficulty is unclear. In this thesis auditory grouping was

an problem in **EXPERIMENT FOUR**; subjects could not set the speech and the $speech_{ipc}$ together in a rhythm setting task. This appeared to be due to a failure to hear the speech and $speech_{ipc}$ sequences as a perceptually coherent stream, about which rhythmic judgements could be made. An issue for further research could thus be whether acoustic inputs need to be segregated perceptually before P-centres are heard, or whether P-centres are determined before the streams are segregated.

This issue in turn sheds light upon experiments in the field of auditory stream segregation which have shown an ambiguous role for rhythm in auditory streaming. McAdams (1990) reported mixed results when investigating the influence of rhythm on sequential auditory organization. Rhythm appeared to affect streaming only when other auditory grouping structures were difficult to form. Handel, Weaver and Lawson (1983) found an effect of rhythm on auditory streaming only if the rhythm was isochronous; they suggested that the auditory system infers the even timing is due to a single source. However in a dynamic rhythm setting task the timing will not be initially isochronous nor the sounds necessarily similar, causing difficulties for rhythm-based grouping of the sequence. If, as seems possible, the acoustic environment needs to be split into streams before P-centres are heard, then rhythm *per se* would not affect stream segregation. Instead stream segregation would affect rhythm perception, as Seton's experiment suggested.

13.7.3 The work on internal clocks, timing illusions and the relevance to P-centre experiments (Lame 1990) was mentioned in the discussion of **EXPERIMENT EIGHT**. The hypothesis is that what sounds occupy the intervals in a dynamic rhythm setting task may affect the perceived rhythm instead of or in parallel with the P-centres of the stimuli, and that P-centres need not therefore be context independent. This relationship could be investigated in an attempt to disambiguate the two affects; this would probably involve attempting

to a find a different method of P-centre determining, perhaps using on of the direct measures described in Chapter Four as well as the dynamic rhythm setting task. This in turn would help disambiguate the variation in the reported effects of vowel onset on dynamic rhythm setting task performance.

13.7.4 In terms of the context independence of P-centres, Seton's 1989 finding that the loudness of a signal and the other signals in a dynamic rhythm setting task affected subjects rhythm setting could be extended. This has direct relevance to the FAIM model, which is affected by the peak amplitude of a signal. Is the slope and location of the amplitude increase sufficient to account for P-centre, or does the relationship vary with different peak amplitude values.

13.8 Summary

The experimental work contained in this thesis was reviewed, with the major findings addressed. The development of the FAIM model of P-centre location, which maps P-centres onto discrete amplitude increases in the frequency bandwidth centre frequency = 578Hz, was described. This is a direct application of a local concept of beat location derived from the musical beat literature (Gordon 1987, Vos and Rasch 1981). P-centers can thus be modelled in a non speech specific, parsimonious manner. The relevance of P-centres to other fields of research, such as synchronization tasks, was described.

Abramson, D., & Lisker, N. (1970). Discrimination along the voicing continuum: Cross language tests. Proceedings of the 6th International Congress of Phonetic Sciences (pp. 569 - 573). Prague: Academic Press.

Allen, G. D. (1972). The Location of rhythmic stress beats in English. Language and Speech, 15, 72-100 and 179 - 195.

Auxiette, C., & Gerard, C. (1992). Perceptual and motor determinants in the synchronization of music and speech. In C. Auxiette, C. Drake & C. Gerard (Ed.), Fourth rhythm workshop: Rhythm perception and production (pp. 59-64). Bourges, France.

Balzano, G. J. (1986). What are musical pitch and timbre? Music Perception, 3(3), 297-314.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequence of tones. Journal of Experimental Psychology, 89(2), 244-249.

Cooper, A. M., Whalen, D. H. & Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. Perception and Psychophysics, 39, 187 - 196.

Cooper, A. M., Whalen, D. H. & Fowler, C. A. (1988). The syllables' rhyme affects its P-center as a unit. Journal of Phonetics, 16(2), 231 - 241.

Cutting, J. E. & Rosner, B. S. (1974). Categories and boundaries in speech and music. Perception and Psychophysics, 16, 564-570.

Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. Perception and Psychophysics, 31(5), 462 - 476.

Darling, A. (1991). Properties and implementation of the GammaTone Filter: A tutorial. In V. Hazan (Ed.), Speech, Hearing and Language: Work in Progress UCL (vol. 5, pp. 43-61). London: Dept. of Phonetics and Linguistics, UCL.

de Boer, E. & Kuyper, P. (1968). Triggered correlation. IEEE Transactions on Biological and Medical Engineering (pp. 169-179).

Delgutte, B. (1980). Representation of speech-like events in the discharge pattern of auditory nerve fibres. Journal of the Acoustical Society of America, 68, 843-857.

Delgutte, B. & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: 1. vowel-like sounds. Journal of the Acoustical Society of America, 73(3), 866-878.

Delgutte, B. & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: 3. Voiceless fricative consonants. Journal of the Acoustical Society of America, 75(3), 887-896.

Delgutte, B. & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: 4. Sounds with consonant-like dynamic characteristics. Journal of the Acoustical Society of America, 75(3), 897-907.

Delgutte, B. (1984). Speech coding in the auditory nerve: II Processing schemes for vowel-like sounds. Journal of the Acoustical Society of America, 75(3), 887-896.

Efron, R. (1970). Effect of stimulus duration on perceptual onset and offset latencies. Perception and Psychophysics, 8, 231-234.

Efron, R. (1970). The minimum duration of a perception. Neuropsychologia, 8, 57-63.

Efron, R. (1970). The relationship between the duration of a stimulus and the duration of a perception. Neuropsychologia, 8, 37-55.

Eling, P. A., Marschall, J. C. & van Galen, G. P. (1980). Perceptual centres for Dutch digits. Acta Psychologica, 46, 95 - 102.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. Perception and Psychophysics, 25(5), 375 - 388.

Fowler, C. A. & Tassinary, L. (1981). Natural measurement criteria for speech: the anisochrony illusion. In J. Long & A. Baddeley (Ed.), Attention and Performance IX. Hillsdale, N. J., Lawrence Erlbaum Associates.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech - cyclic production of vowels in monosyllablic stress feet. Journal of Experimental Psychology - General, 112(3), 386 - 412.

Fowler, C. A. & Tassinary, L. G. (1986). Perception of syllable timing by prebabbling infants. Journal of the Acoustical Society of America, 79(3), 814 - 825.

Fowler, C. A., Whalen, D. H. & Cooper, A. M. (1988). Perceived timing is produced timing: A reply to Howell. Perception and Psychophysics, 43, 93 - 98.

Fox, R. A. & Lehiste, I. (1987). The effect of vowel quality variations on stress-beat location. Journal of Phonetics, 15(1), 1 - 13.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), The Psychology of music (pp. 149-181). New York: Academic Press.

Gerstman, L. J. (1957). Perceptual dimensions for the frication portion of certain speech sounds. Ph.D., New York.

Gjaevenes, P. & Rimstad, T. (1972). The influence of rise time on the loudness of sound pulses. Journal of the Acoustical Society of America, 51, 1233 - 1239.

Gordon, J. W. (1987). The Perceptual attack time of musical tones. Journal of the Acoustical Society of America, 82(1), 88 - 105.

Handel, S., Weaver, M. S. & Lawson, G. (1983). Effect of rhythmic grouping on stream segregation. Journal of Experimental Psychology: Human Perception and Performance, 9(4), 637-651.

Hirsh, I. J. (1959). Auditory perception of temporal order. Journal of the Acoustical Society of America, 31, 759-767.

Hoequist, C. E. (1983). The P-center and rhythm categories. Language and Speech, 26(4), 367-376.

Holdsworth, J., Nimmo-Smith, I., Patterson, R. & Rice, P. (1988). Implementing a GammaTone filterbank, MRC Applied Psychology Unit, Cambridge, England.

Howell, P. (1984). An acoustic determinant of perceived and produced isochrony. In M. P. R. Van den Broecken & A. Cohen (Ed.), 10th International Congress of Phonetic Sciences (pp. 429 - 433), Dordrecht, Holland: Foris.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: 1. Perception and Psychophysics, 43, 90-93.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: 2. Perception and Psychophysics, 43, 99.

Johannesma, P. I. M. (1972). The pre-response stimulus ensemble of neurones in the cochlear nucleus. Symposium on Hearing Theory (pp. 58-69). IPO, Eindhoven, The Netherlands.

Kewly-Port, D. & Pisoni, D. B. (1984). Identif1 3 491 ncategorical? Journal of the Acoustical Society of America, 75, 1168-1176.

Kuhl, P. K. & Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63(3), 905-917.

Ladefoged, P. (1982). A Course in Phonetics. San Diego: Harcourt Brace Jovanovich.

Lame, G. D. (1990). Timing perception and control: A new internal clock model and modality-specific phase resetting, Ph.D., University of Texas at Austin.

Lee, C. S. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West & I. Cross (Ed.), Representing musical structure (vol. 5, pp. 59-128). London: Academic Press.

Lehiste, I. (1977). Isochrony reconsidered. Journal of Phonetics, 5, 253 - 263.

Lerdahl, F. & Jackendoff, R. A. (1983). A generative theory of tonal music. Cambridge MA: MIT Press.

Lerdhal, F. & Jackendoff, R. A. (1983). A Generative Theory of Tonal Music. Cambridge, M.A., MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Marcus, S. (1976). Perceptual Centres. Ph.D., University of Cambridge.

Marcus, S. M. (1981). Acoustic determinants of perceptual centre (P-center) location. Perception and Psychophysics, 30, 247 - 256.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behaviour. Psychological Review, 7(6), 487-509.

McAdams, S. (1990). Influences of rhythm on sequential auditory organization. 2nd International Conference on Music and the Cognitive Scinces Cambridge UK.

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. Journal of the Acoustical Society of America, 58, 880-883.

Miller, M. (1984). On the perception of rhythm. Journal of Phonetics, 38, 159-180.

Morton, J., Marcus, S. M. & Frankish, C. (1976). Perceptual centres (P-centers). Psychological Review, 83, 405 - 408.

Pastore, R. E., Harris, L. B. & Kaplan, J. K. (1982). Temporal order identification: Some parameter dependencies. Journal of the Acoustical Society of America, 71(2), 430 - 436.

Pastore, R. E. (1983). Temporal order judgment of auditory stimulus offset. Perception and Psychophysics, 33(1), 54-62.

Patterson, R. & Moore, B. C. J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In B. C. J. Moore (Ed.), Frequency Selectivity in Hearing (pp. 123-177). London: Academic.

Patterson, R. D., Holdsworth, J. & Allerhand, M. (1991). Auditory models as preprocessors for speech recognition. In B. Schouten (Ed.), The Psychophysics of speech perception 2 Utrecht.

Pompino-Marschall, B. & . (1989). On the psychoacoustic nature of the P-center phenomenon. Journal of Phonetics, 17(3), 175 - 192.

Pompino-Marschall, B. & . (1991). The syllable as a prosodic unit and the so-called P-center effect (29). Instituts fur Phonetik und Sprachliche Kommunikation der Universitat Munchen (FIPKM).

Prinz, W. (1992). Distal focussing in action control. In C. Auxiette, C. Drake, & C. Gerard (Ed.), Fourth rhythm workshop: Rhythm perception and production (pp. 65-71). Bourges, France:

Rapp, K. (1971). A Study of syllable timing (1). Speech Transmission Laboratories - Quarterly Papers on Speech Research, pp. 14-19.

Rasch, R. (1979). Synchronization in performed ensemble music. Acustica, 43, 121 - 131.

Rosen, S. M. & Howell, P. (1981). Plucks and bows are not categorically perceived. Perception and Psychophysics, 30, 156-161.

Schutte, H. (1977). Bestimmung der subjektiven Erreigniszeitpunkte aufeinanderfolgender. Ph.D., Technical University, Munich.

Schutte, H. (1978). Ein Funktionsschema fur die Wahrnehmung eines gleichmassigen. Biological Cybernetics, 29, 49-55.

Schutte, H. (1978). Subjekiv gleichmassinssen Rhythmus: Ein Beitrag zur reitlichen Wahrnehmung von Schallereignissen. Acustica, 14, 197-206.

Scott, D. R., Isard, S. D. & de Boysson-Bardies, B. (1985). Perceptual isochrony in English and French. Journal of Phonetics, 13, 155 - 162.

Selkirk, E. O. (1984). Phonology and syntax: the relation between sound and structure. Cambridge, MA: MIT press.

Semjen, A., Schulze, H. & Vorberg, D. (1992). Temporal control in the coordination between repetitive tapping and periodic external stimuli. In C. Auxiette, C. Drake & C. Gerard (Ed.), Fourth rhythm workshop: rhythm perception and production (pp. 73-78). Bourges, France:

Seton, J. C. (1989). A psychophysical investigation of auditory rhythmic beat perception. Ph.D., University of York.

Smurzynski, J. & Houtsma, A. J. M. (1989). Auditory discrimination of tone-pulse onsets. Perception and Psychophysics, 45(1), 2-9.

Suga, N. (1971). Responses of Inferior Collicular neurones of bats to tone bursts with different rise times. Journal of Physiology, 217, 159 - 177.

Terhart, E. (1978). Psychoacoustic evaluation of musical sounds. Perception and Psychophysics, 23(6), 483 - 492.

Tuller, B. & Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. Perception and Psychophysics, 27, 277 - 283.

Tuller, B. & Fowler, C. A. (1981). The Contribution of amplitude to the perception of isochrony (SR - 65). Haskins Laboratories.

Tye-Murray, N., Zimmermann, G. N. & Folkins, J. (1987). Movement timing in deaf and hearing speakers: Comparison of phonetically heterogeneous syllable strings. Journal of Speech and Hearing Research, 30(3), 411-417.

van der Broeke, M. P. R. & van Heuven, V. J. (1983). Effect and Artifact in the auditory discrimination of rise and decay time: Speech and nonspeech. Perception and Psychophysics, 33(4), 305-313.

van Heuven, J. V. & van de Broeke, M. P. R. (1979). Auditory discrimination of rise and decay times in tone and noise bursts. Journal of the Acoustical Society of America, 66(5), 1308-1315.

Vos, J. & Rasch, R. (1981). The Perceptual onset of musical tones. Perception and Psychophysics, 29(4), 323-335.

Whalen, D. H., Cooper, A. M. & Fowler, C. A. (1989). P-center judgments are generally insensitive to the instructions given. Phonetica , 46(4), 197 - 203.

Appendix: code to determine Marcus Model predictions

1. Filtering

```
# shell script to determine the vowel onset of a speech signal
# defined as peak increment in mid-band spectral energy

if ($#argv == 0) then
        echo "no sfs file stated"
        exit 1
endif

set D = 'pwd'
echo filtering $argv[1]....
genfilt -l1500 -h500 $D/$argv[1]

echo working out time of peak increment
sml -i1.02 /users/sophie/model/marcus_csh.sml $D/$argv[1]
```

---

2. Vowel onset detection

```
/* program to go through a speech file and integrate energy over 10ms*/
/*moving forward in half window sized steps*/
/* designed to run under shell script */
/* echoes the time of peak increment in energy */

main {
        var st,et,xt,yt
        var t,bill,ben
        var t0,t1,samp
        var time, bigger,lastbill
        var window

        st = 0
        et = 0.500
        window = 0.01

        t0 = next(SP,-1)
        t1 = next(SP,t0)
        print "sampling rate = ",1.0/(t1-t0),"Hz\n"
        samp = (1.0/(t1-t0))
```

```
        yt = (st + window)
        t = next(SP,(st-1))
        time = st
        bill = 0
        bigger = 0
        while (t<=et) {
                        lastbill=bill
                        bill = 0
                        while (t<yt) {
                                ben = sp(t)
                                if (ben<0) {
                                        ben = ben*(-1)
                                }
                                bill = bill+ben
                                t = next(SP,t)
                        }

                if ((bill-lastbill)>bigger) {
                        bigger = (bill-lastbill)
                        time = st
                }
                st = st + (window/2.0)
                yt = yt + (window/2.0)
        }
        print "peak increment in midband spectral energy =",time,"\n"
        print "and it's size is :", bigger,"\n"
        if (t > et){
                abort( " END\n")
        }
}
```

---

3. P-centre calculation

```
/* program to calculte the marcus model p center predictions */
main {
        var vot, dur, rest

        print "enter the duration of the sound :"
        input dur
        print "enter the vowel onset time :"
```

```
        input vot
        rest = (dur - vot)
        print "The Marcus model P center prediction =
",((vot*0.65)+(rest*0.25))," +K\n"
}
```

Appendix: code to determine the syllabic centre of gravity of a speech signal (Howell 1988a)

```
/* program to find center of gravity of a speech signal */

main {
        var st,et
        var t,bill,ben,weed

        print "enter start time in seconds : "
        input st
        print "enter end time in seconds : "
        input et
        bill = 0
        weed = 0
        t = next(SP,(st-1))
        while (t<=et) {
                if ((t>st) && (t<et)) {
                        ben = sp(t)
                        if (ben<0) {
                                ben = ben*(-1)
                        }
                        bill = bill+ben
                }
                t = next(SP,t)
        }
        print "integrated samples = ",bill,"\n"
        t = next(SP,(st-1))
        while (weed<(bill/2)) {
                if ((t>st) && (t<et)) {
                        ben =sp(t)
                        if (ben<0) {
                                ben = ben*(-1)
                        }
                        weed = weed+ben
                }
                t = next(SP,t)
        }
        print "time at 1/2 energy = ",(t-0.001)," ms\n"
}
```

Appendix: code to calculate the Vos and Rasch Perceived Onset of Signals (Vos and Rasch 1981)

```
/* to calculate Vos and Rasch Perceptual Onset threshold */
/* equal to peak intensity in dB minus 15dB */


file out
main {
        var     t,st,et,mt
        var     f                       /* raw amplitude */
        var     volt                    /* amplitude expressed in volts */
        var     volref                  /* amplitude divided by reference voltage */
        var     deeb                    /* intensity in dB */
        var     logx, logten
        var     max, thresh
        string  dat_file                /* outfile */

        print "enter output data file : "
        input dat_file
        openout (out, dat_file)
        print "enter start time:"
        input st
        print "enter end time:"
        input et
        f = 0
        volt = 0
        volref = 0
        deeb = 0
        max = 0
        t = next(SP,st)
        while (t < et) {
                f = sp(t)
                f = abs(f)
                volt = (5*f)/2047 /* converting amplitude into voltage units */

/* change the reference voltage by altering the value of the next divisor */
/* in volts that is! */

                volref = (volt/0.000001)
                volref = (volref*volref)
                logx = log(volref)
                logten = log(10)
                deeb = 10*(logx/logten)
```

```
        if (deeb > max) {           /* find maximum */
                max = deeb
                mt = t              /* time of maximum */

        }
        t = next(SP,t)
    }
    thresh = (max - 15.0)
    t = next (SP, st)               /* find threshold */
    while (t < mt) {                /* while t < t(max) */
        f = sp(t)
        f = abs(f)
        volt = (5*f)/2047
        volref = (volt /0.000001)
        volref = (volref*volref)
        logx = log(volref)
        deeb = (10*(logx/logten))
        if (deeb >= thresh)  {
                print # out " threshold \t", thresh, "at time \t ", t, "\n"
        }
        t = next(SP,t)

    }

    print "its over:\n"
}
```

Appendix: P-centre fitting algorithm - Marcus 1976

```
C program to perform p-center analysis on a file
C of experimental data output by a member of the
C ISOWAV family (v0.07 or later)
C input file format:
C subject name (10a2)
C testing date (10a2)
C condition id (10a2)
C sequence filename (3a2)
C NSTIM: the number of stimuli (i5)
C NSTIM x waveform filenames (NSTIM rows of 3a2 data)
C an NSTIM x NSTIM data matrix (each row with NSTIM x i5 entries)
C routines called:
C                CENTRE,INVMAT,OUTPUT
C CENTRE performs a best p-center fit to the data matrix
C INVMAT inverts the "number of observations" matrix in CENTRE
C OUTPUT writes out the results of the analysis on the vdu
C all variables are dimensioned and passed to subroutines
C via the COMMON block: 18 lines of comment

      COMMON IY(10,10),N(10,10),IDUMMY(10),MATLIN(10)
      COMMON B(10),NDF,FIT,INAME(10),IDATE(10),IDENT(10)
      COMMON NSTIM,ISEQUE(3),IDNAMS(10,10),INFILE(10)
      COMMON
LOGNIT(10,10),A(10,10),D(10),IPVT(10),PVT(10),IND(10,2)
      REAL PCPOOL(10,20),PAVRGE(10)
C initialise all arrays etc. with 0's & blanks as appropriate.....
      DO 2111 I=1,10
      DO 2111 J=1,10
      IY(I,J)=0
      N(I,J)=0
      B(I)=0.
      INAME(I)=0
      IDATE(I)=0
      IDENT(I)=0
      LOGNIT(I,J)=0
      A(I,J)=0.
      D(I)=0.
      IPVT(I)=0
      PVT(I)=0.
      PAVRGE(I)=0.
2111  CONTINUE
      DO 2112 I=1,10
```

```
      DO 2112 J=1,3
      ISEQUE(J)=0
      IDNAMS(I,J)=0
      INFILE(J)=0
2112  CONTINUE
      DO 2113 I=1,10
      DO 2113 J=1,2
      IND(I,J)=0
 2113 CONTINUE
      DO 2114 I=1,10
      DO 2114 J=1,20
      PCPOOL(I,J)=0.
 2114 CONTINUE
      WRITE(6,10)
10    FORMAT(' CPOOL: p-center analysis program version 0.01'///)
C ask if data is to be pooled... line 58
      IPOOL=0
      NTIMES=1
C WHAT IS SIG OF -1 AS OPP 1?
      WRITE(6,12)
12    FORMAT(1H ,' type 1 and <CR> to pool data, else hit <CR>',/)
      INCHAR=0
      READ(5,13)INCHAR
13    FORMAT(I1)
      IF(INCHAR-1)500,14,500
C come here if we are pooling....
14    IPOOL=1
C and find out how many subjects to pool over....
C NEXT WRITE WAS -1
      WRITE(6,141)
141   FORMAT(1H ,' how many subjects are there?',/)
      READ(5,142)NTIMES
142   FORMAT(I5)
      IF(NTIMES)43,43,44
44    IF(NTIMES-20)45,45,43
43    WRITE(6,46)
46    FORMAT(1H1,' number of subjects not in range 1 - 20 ')
      GOTO 14
45    CONTINUE
500   CONTINUE
C for each subject in turn....
      DO 2000 NREPS=1,NTIMES
C get the data filename.....
5     CONTINUE
C -1
```

```
C       WRITE(6,20)
C20     FORMAT(1H ' enter the data filename.....')
C       READ(5,30)INFILE
C30     FORMAT(3A2)
C ALREADY 6 nb sophie at present works for fixed file name CALLLED
INFILE
C       WRITE(6,30)INFILE
C here if the file is found ok.....
C open the disc file for reading....
60      CONTINUE
C*******

OPEN(4,FILE='/users/sophie/pcenstuff/infile',ERR=2999,FORM='FORMATTE
D')
C get the name/date/condition data from the input file...
C       READ(4,80)INAME
C       READ(4,80)IDATE
C       READ(4,80)IDENT
80      FORMAT(A6)
C read in the sequence file name
C       READ(4,85)ISEQUE
C       WRITE(6,851)ISEQUE
85      FORMAT(A6)
851     FORMAT(1H ,'SEQUENCE IS',A6)
C get the number of different stimuli used in this experimental run...
        READ(4,90)NSTIM
90      FORMAT(I5)
        WRITE(6,901)NSTIM
901     FORMAT(1H ,'NUMBER OF STIM=',I5)
C read in the NSTIM * NSTIM matrix of response data.....

READ(4,202)((IY(LOOP1,LOOP2),LOOP1=1,NSTIM),LOOP2=1,NSTIM)
 202    FORMAT(I5)
        DO 210 LOOP1=1,NSTIM
        DO 210 LOOP2=1,NSTIM
        N(LOOP1,LOOP2)=1
 210    CONTINUE
C close the read channel....
        CLOSE(4)
C work out the mean and sd for all filled cells....
C initialise the sums etc....
        DSUM=0.0
        DSSQ=0.0
        DMEAN=0.0
        DSD=0.0
```

```
      DVZOR=(NSTIM*NSTIM)-NSTIM
C and compute for the whole data matrix....
      DO 1000 LOOP1=1,NSTIM
          DO 1001 LOOP2=1,NSTIM
C ignore the diagonals...
                IF(LOOP1-LOOP2)1002,1001,1002
1002            DSUM=DSUM+FLOAT(IY(LOOP1,LOOP2))
1001      CONTINUE
1000  CONTINUE
C work out the average....
      DMEAN=DSUM/DVZOR
C and now do the standard deviation....
      DO 1100 LOOP1=1,NSTIM
          DO 1101 LOOP2=1,NSTIM
C again, ignore the diagonals....
                IF(LOOP1-LOOP2)1102,1101,1102
1102            DEVIAT=DMEAN-(FLOAT(IY(LOOP1,LOOP2)))
                DSSQ=DSSQ+(DEVIAT*DEVIAT)
1101      CONTINUE
1100  CONTINUE
C work out variance and standard deviation...
      DVAR=DSSQ/(DVZOR-1)
      DSD=SQRT(DVAR)
C and say what they are.....
C ALREADY 5
      WRITE(6,1200)DMEAN,DSD
1200   FORMAT(//' mean offset = ',F8.3,' milliseconds ',2X/
   1,' standard deviation = ',F8.3,' milliseconds '//)
C now go off and do the P-center analysis....
      WRITE(6,1111)
1111  FORMAT(1H ,'GOING INTO CENTRE',1H )
      WRITE(6,556)((IY(IO,IU),IO=1,10),IU=1,10)
      WRITE(6,556)((N(IO,IU),IO=1,10),IU=1,10)
 555  FORMAT(10F6.2)
 556  FORMAT(10I6)
      CALL CENTRE
      WRITE(6,2222)
2222  FORMAT(1H ,'OUT OF CENTRE',1H )
C and then write the results out...
      WRITE(6,333)NDF
      WRITE(6,334)(B(MLK),MLK=1,10)
 333  FORMAT(I6)
 334  FORMAT(F10.2)
      CALL OUTPUT
C are we pooling subjects?....
```

```
      IF(IPOOL)2001,2000,2001
C if so, there's some extra maths to do!....
2001  CONTINUE
      DO 2002 INDEX=1,NSTIM
      PCPOOL(INDEX,NREPS)=B(INDEX)
2002  CONTINUE
2000  CONTINUE
C now pool the p-centres if required.....
      IF(IPOOL)2100,2999,2100
2100  CONTINUE
      DO 2200 JSTIM=1,NSTIM
          PTOTAL=0.0
          DO 2300 JSUBJ=1,NTIMES
          PTOTAL=PTOTAL+PCPOOL(JSTIM,JSUBJ)
2300      CONTINUE
          PAVRGE(JSTIM)=PTOTAL/FLOAT(NTIMES)
2200  CONTINUE
C now write the pooled data out.....LINE 186
      DO 2400 L=1,5,4
      WRITE(6,2410)
2410  FORMAT(/////' here are the pooled p-centres.....'///)
      DO 2420 LAVA=1,NSTIM
      IL1=IDNAMS(LAVA,1)
      IL2=IDNAMS(LAVA,2)
      IL3=IDNAMS(LAVA,3)
      WRITE(6,2422)IL1,IL2,IL3,PAVRGE(LAVA)
2422  FORMAT(1H ,' stimulus   ',3A2,'   pooled p-centre is :',F8.3,/)
2420  CONTINUE
2400  CONTINUE
C and exit...
2999  STOP
      END
C*************************************************************
C*************************************************************
      SUBROUTINE CENTRE
C least square P-centre fit to observed data matrix IY(I,J)
C    each data entry is the average of N(I,J) observations
C matrices Y and N are assumed to be 10 x 10,
C not all the cells need be filled
C returns NDF degrees of freedom for residual FIT
C and best fit p-centre vector B(10)
C M.I.Nimmo-Smith, Applied Psychology Unit, Cambridge, England,
C 1975. This copy derived from a printed Fortran 4 source from
C    Stephen Marcus, IPO
C modified to run as fortran-2 by JCS 18/4/83
```

```
C AND MODIFIED TO RUN ON MASSCOMP PH 23/4/92
      COMMON IY(10,10),N(10,10),IDUMMY(10),MATLIN(10)
      COMMON B(10),NDF,FIT,INAME(10),IDATE(10),IDENT(10)
      COMMON NSTIM,ISEQUE(3),IDNAMS(10,10),INFILE(10)
      COMMON
LOGNIT(10,10),A(10,10),D(10),IPVT(10),PVT(10),IND(10,2)
C
      NDF=0
      DO 40 I=1,NSTIM
      IY(I,I)=0
      N(I,I)=0
 40   CONTINUE
C
C FOLD MATRIX Y
      NM1=NSTIM-1
      DO 50 I=1,NM1
      I1=I+1
      DO 50 J=I1,NSTIM
      M=N(I,J)+N(J,I)
      NY=IY(I,J)*N(I,J)-IY(J,I)*N(J,I)
      IY(I,J)=0
C TAKE MEAN OF DIAGONAL CELLS
      N(I,J)=M
      N(J,I)=0
      IF(M)50,50,55
55    CONTINUE
      IY(I,J)=NY/M
      NDF=NDF+1
50    CONTINUE
      DO 100 I=1,NSTIM
      B(I)=0
      D(I)=0
      DO 100 J=1,NSTIM
100   A(I,J)=1
      DO 150 I=1,NSTIM
      DO 150 J=1,NSTIM
      X=N(I,J)
      A(I,I)=A(I,I)+X
      A(J,J)=A(J,J)+X
      A(I,J)=A(I,J)-X
      A(J,I)=A(J,I)-X
      ISG=J-I
      X=0
      IF(ISG)160,170,160
160   X=ISIGN(1,ISG)*N(I,J)*IY(I,J)
```

```
170   D(I)=D(I)+X
      D(J)=D(J)-X
150   CQNTINUE
C ALLOW FOR EMPTY CELLS
      NCELL=NSTIM
      DO 200 I=1,NSTIM
      DO 210 J=1,NSTIM
      IF(N(I,J)+N(J,I))210,210,200
210   CONTINUE
      DO 220 J=1,NSTIM
      A(I,J)=0
220   A(J,I)=0
      A(I,I)=1
      NCELL=NCELL-1
200   CONTINUE
      CALL INVMAT
      DO 400 I=1,NSTIM
      B(I)=0
      DO 300 J=1,NSTIM
300   B(I)=B(I)+A(I,J)*D(J)
C B VECTOR IS LEAST SQUARE P-CENTRE FIT
400   CONTINUE
C CALCULATE RESIDUAL = VARIANCE PER DATA POINT FROM
P-CENTRE FIT
      FIT=0
      DO 450 I=1,NSTIM
      DO 450 J=1,NSTIM
      X=IY(I,J)
      XN=N(I,J)
450   FIT=FIT+XN*(X-B(I)+B(J))**2
      NDF=NDF-NCELL
      FIT=FIT/FLOAT(NDF)
      RETURN
      END
C*****************************************************************
      SUBROUTINE INVMAT
C INVERT MATRIX A BY GAUSS-JORDAN METHOD
      COMMON IY(10,10),N(10,10),IDUMMY(10),MATLIN(10)
      COMMON B(10),NDF,FIT,INAME(10),IDATE(10),IDENT(10)
      COMMON NSTIM,ISEQUE(3),IDNAMS(10,10),INFILE(10)
      COMMON
LOGNIT(10,10),A(10,10),D(10),IPVT(10),PVT(10),IND(10,2)
      DET=1
      DO 1 J=1,NSTIM
      IPVT(J)=0
```

```
1       CONTINUE
        DO 10 I=1,NSTIM
        AMAX=0.0
        DO 5 J=1,NSTIM
        IF(IPVT(J)-1)2,51,2
2       CONTINUE
        DO 52 K=1,NSTIM
        IF(IPVT(K)-1)3,53,200
3       IF(ABS(AMAX)-ABS(A(J,K)))4,53,53
4       IROW=J
        ICOL=K
        AMAX=A(J,K)
53      CONTINUE
52      CONTINUE
51      CONTINUE
5       CONTINUE
        IPVT(ICOL)=IPVT(ICOL)+1
        IF(IROW-ICOL)6,8,6
6       DET=-DET
        DO 7 L=1,NSTIM
        SWAP=A(IROW,L)
        A(IROW,L)=A(ICOL,L)
7       A(ICOL,L)=SWAP
8       IND(I,1)=IROW
        IND(I,2)=ICOL
        PVT(I)=A(ICOL,ICOL)
        DET=DET*PVT(I)
        A(ICOL,ICOL)=1
        DO 9 L=1,NSTIM
9       A(ICOL,L)=A(ICOL,L)/PVT(I)
        DO 10 L1=1,NSTIM
        IF(L1-ICOL)11,10,11
11      SWAP=A(L1,ICOL)
        A(L1,ICOL)=0
        DO 12 L=1,NSTIM
12      A(L1,L)=A(L1,L)-A(ICOL,L)*SWAP
10      CONTINUE
        DO 20 II=1,NSTIM
        L=NSTIM+1-II
        IF(IND(L,1)-IND(L,2))13,200,13
13      IROW=IND(L,1)
        ICOL=IND(L,2)
        DO 21 K=1,NSTIM
        SWAP=A(K,IROW)
        A(K,IROW)=A(K,ICOL)
```

```
      A(K,ICOL)=SWAP
21    CONTINUE
200   CONTINUE
20    CONTINUE
C A = INVERSE (A)
C DET = DET(A)
      RETURN
      END

C*************************************************************
      SUBROUTINE OUTPUT
C routine to write out the results of the p-center analysis
      COMMON IY(10,10),N(10,10),IDUMMY(10),MATLIN(10)
      COMMON B(10),NDF,FIT,INAME(10),IDATE(10),IDENT(10)
      COMMON NSTIM,ISEQUE(3),IDNAMS(10,10),INFILE(10)
      COMMON
LOGNIT(10,10),A(10,10),D(10),IPVT(10),PVT(10),IND(10,2)
C     LOGNIT=1
5     CONTINUE
      WRITE(6,10)
10    FORMAT(' results of p-center analysis.....'//)
      WRITE(6,20)INAME
      WRITE(6,21)IDATE
      WRITE(6,22)IDENT
20    FORMAT(5X,'subject name.....',10A2)
21    FORMAT(5X,'testing date.....',10A2)
22    FORMAT(5X,'condition code....',10A2)
      WRITE(6,23)ISEQUE
23    FORMAT(1H ,' sequence filename was.....',3A2)
      WRITE(6,25)
25    FORMAT('  stimulus        p-center (milliseconds)'/)
      DO 30 LOOP1=1,NSTIM
         DO 40 LOOP2=1,NSTIM
40       IDUMMY(LOOP2)=IDNAMS(LOOP1,LOOP2)
         WRITE(6,50)IDUMMY,B(LOOP1)
50       FORMAT(4X,A6,14X,F10.2)
C     WRITE(6,50)(B(LOOP1),LOOP1=1,NSTIM)
C50   FORMAT(1H ,24X,F10.2)
30    CONTINUE
      WRITE(6,60)FIT
60    FORMAT(1H ,' fit.....',X,F10.2)
      WRITE(6,70)NDF
70    FORMAT(1H ,' with  ',I5,'  degrees of freedom')
C     IF(LOGNIT-5)1000,1001,1000
C1000       LOGNIT=5
```

```
C       GOTO 5
1001    CONTINUE
        RETURN
        END
```

Code to control dynamic rhythm setting experiments - Stevie Sackin 1991

```
#define PROGNAME "Rhythm"
#define PROGVERS "4.0"
char *progname=PROGNAME;

#include "sfs.h"
#include <time.h>
#include <stdio.h>
#include <math.h>          /* for random playback routine */

#define BA 0x218
#define MAXSTRING 100
#define p_a 0x300
#define p_b 0x301
#define p_c 0x302
#define control_reg 0x303
#define CW 0x98

/* global variables */

struct item_header spitem[20];
short *sp[20];

void extern zdelay();
void extern playback();

main(argc, argv)
int argc;
char *argv[];
{
        int aaint, abint, resultarray[100], ptr, nl, aa, ab;
        FILE *fp, *fpin;
        int abstart;
        int d, m, y, dow, i, errflag=0;
        char set[10], x[10], temp[10];
        char file[30];
        float sf;
        unsigned f, c0, c1;
        char control[3];

        if (argc<2) {
                printf("usage: %s <filename>\n", PROGNAME);
```

```
        exit(1);
}

if ((fp = fopen(argv[1], "a+")) == NULL) {
        printf("can't open %s", argv[1]);
        exit(1);
}

if (argc > 2) {
        if ((fpin = fopen(argv[2], "r")) == NULL) {
                printf("can't open %s", argv[2]);
                exit(1);
        }
}

m=0;
initdt2811(&m);

/* Print date to outfile */
dosdat(&m, &d, &y, &dow);
fprintf(fp, "\n\nToday is: %d-%d-%d", d,m,y);
time(&d);
srand(d);

/* write stimuli etc. to outfile */
fprintf(fp, "\n2 stimuli used from file: ");
fscanf(fpin, "%s %s %s %d %d", file, temp, set, &aa, &aaint);
fprintf(fp, "%s. These are;\n\n%s\t%s", file, temp, set);
getitem(file, SP_TYPE, temp, &spitem[0], &sp[0]);
getitem(file, SP_TYPE, set, &spitem[1], &sp[1]);

/* set scale factor for a/d conversions */
sf = 0.1; /* 200/2000 */

outpw(control_reg, CW); /* set up 8255 for lcd and start/stop */
initlcd();

/* get DT2811 clock code */
f = (int) (1.0/spitem[0].frameduration);
c0 = dac_nearestfreq(&f);
f = (int) (1.0/spitem[1].frameduration);
c1 = dac_nearestfreq(&f);

do {
        lcd("Centre nob.");
```

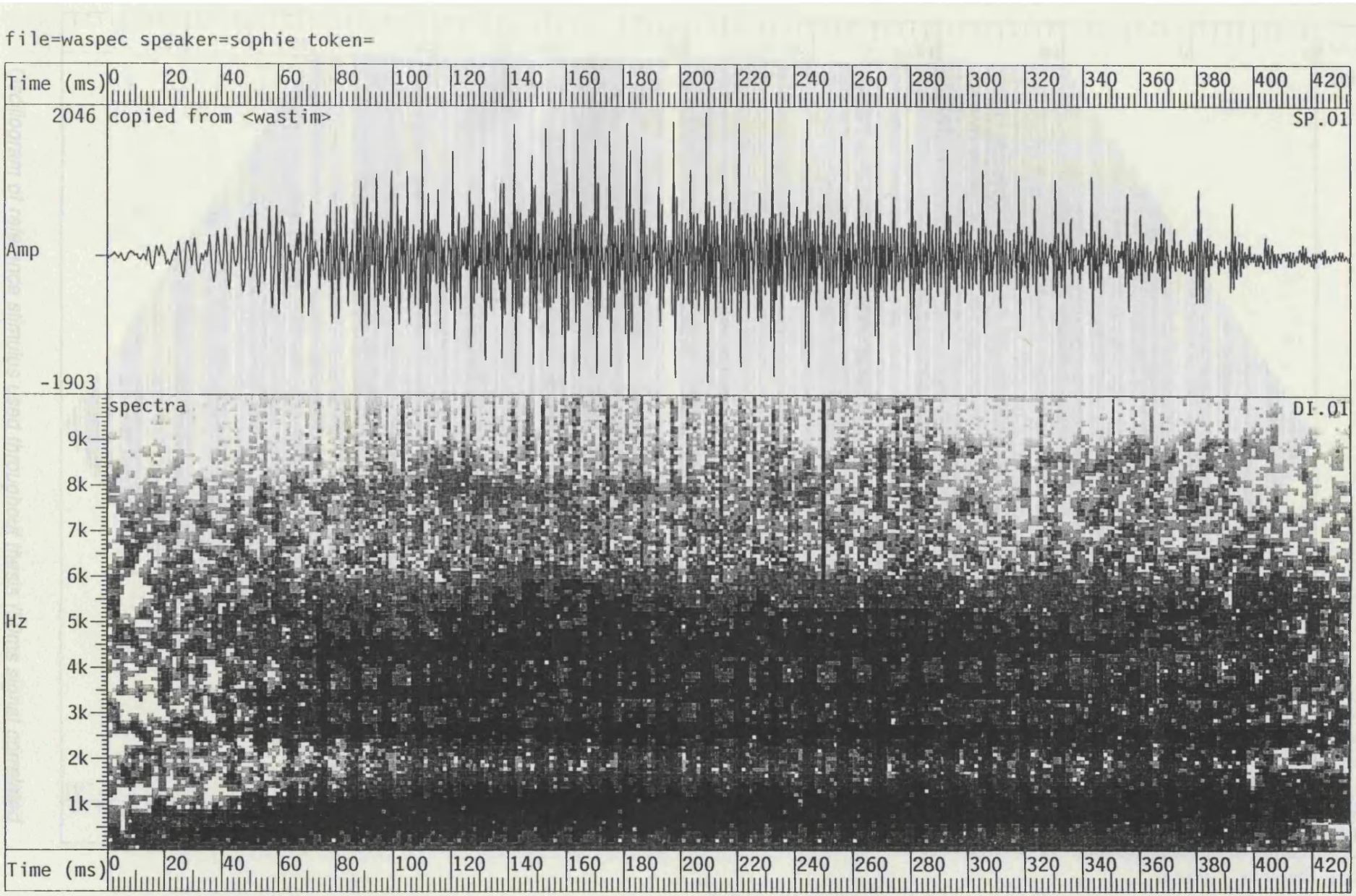
```
                zdelay(1000);

                /* randomise start intervals and write to outfile */
                makerandomstart(aaint, &abint, sf);
                fprintf(fp, "\n\n\nStarting Intervals -\t\tA-A = %dms\tA-B =
%dms", aaint, abint);
                fprintf(fp, "\nNumber of Repetitions -\t\tA-A = %d\t\t", aa);


printf("\n\n*************************************************************************\n\
n'A'-'A' interval = %d,\nStarting 'A'-'B' interval = %d,\n\n", aaint, abint);

                /* subtract length of sounds from intervals */
                aaint -= (int)(((spitem[1].frameduration *
(float)spitem[1].numframes) + (spitem[0].frameduration *
(float)spitem[0].numframes)) * 1000.0);
                abint -= (int)((spitem[0].frameduration *
(float)spitem[0].numframes) * 1000.0);
                abstart = abint;

                /* error checking */
                if ((aaint < 0) || (abint < 0) || (abint > aaint)) {
                        printf("\n**** WARNING - Intervals shorter than sounds
or A-B interval too long ****\n");
                        exit(1);
                }

                m=0; nl=1; ab=0;
                initdt2811(&m);

                ptr = errflag = 0;
                lcd("Hit Keyboard to start");
                zdelay(500);
                while (nl) (nl=inpw(port_c) & 0x10);
                d=1;
                lcd("Hit Keyboard to stop");
                zdelay(500);
                for (i=0; i<aa; i++) {  /* start by playing A alone */
                        outpw(BA+1, 0x00);
                        dac32(BA, sp[0], spitem[0].numframes, 1, c0, &nl,
port_c);
                        zdelay(aaint + (int)((spitem[1].frameduration * (float)
spitem[1].numframes) * 1000.0));
                }
                nl=1;
```

```
                do {  /* introduce B and check responses */
                        outpw(BA+1, 0x00);
                        dac32(BA, sp[0], spitem[0].numframes, 1, c0, &nl,
port_c);
                        initdt2811(&d);
                        checkinp(&abint, sf, resultarray, &ptr, aaint, m, abstart);
/* poll input */
                        zdelay(abint, &nl);
                        outpw(BA+1, 0x00);
                        dac32(BA, sp[1], spitem[1].numframes, 1, c1, &nl,
port_c);
                        initdt2811(&d);
                        checkinp(&abint, sf, resultarray, &ptr, aaint, m, abstart);
/* poll input */
                        zdelay(aaint - abint, &nl);
                        ab++;
                } while (nl);
                lcd("End of session.");
                                aaint += (int)(((spitem[1].frameduration *
(float)spitem[1].numframes) + (spitem[0].frameduration *
(float)spitem[0].numframes)) * 1000.0);
                printresults(resultarray, &ptr, fp, ab);
        } while (errflag != 0);
        fprintf(fp, "\n\n****************************** End of Session
**************************\n\n\n\n");
        lcd("Turn off nob");
        printf("\n\n\n******  Results written to ** %s **\n\n\n", argv[1]);
        fclose(fp);
} /* end of main */

initdt2811(x)
int *x;
{
        int lb, hb;
        int csr, mask, a, flag=0;

        outpw(BA, 0x00);    /* initialise board */
        zdelay(1);
        lb = inpw(BA+2);        /* read low byte */
        hb = inpw(BA+3);    /* read high byte */
        outpw(BA, 0x11);    /* set mode 1 and clear ADDERR */
        outpw(BA+7, 0x22);          /* set base freq = 1.5 kHz */
        outpw(BA+1, 0x00);          /* set gain = 1, ch = 1. Start
                                continuous conversions */
```

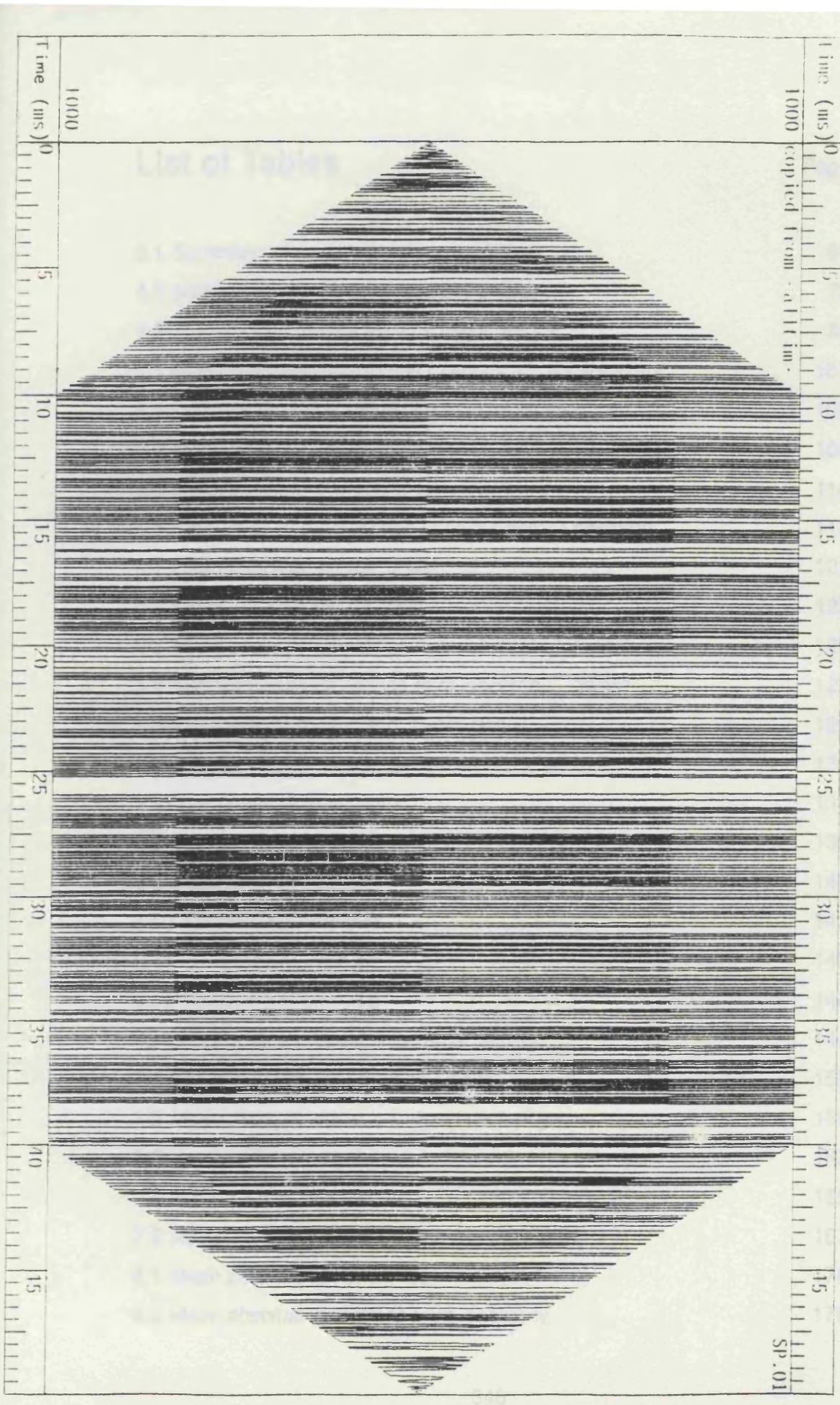
```
        if (*x == 0) do {
                csr = inpw(BA);                 /* read csr */
                mask = 0xc0;                    /* mask csr bits 5-0 */
                a = csr & mask;
                if ((a == 0xc0) || (a == 0x40)) {   /* A/D ERR set? */
                        printf("******  A/D ERROR  ******\n\n");
                        outp(BA, 0x10);
                /*      exit(1); */
                }
                if (a == 0x80) {                /* A/D DONE set? */
                        csr = inp(BA+2);        /* read low and high bytes */
                        a = inp(BA+3);
                        *x = (a << 8) + csr;
                        flag++;
                }
        } while (flag == 0);
}

checkinp(delay, sf, resultarray, ptr, aaint, sv, abs)
int *delay, resultarray[], *ptr, aaint, sv, abs;
float sf;
{
        int csr, mask, a, bog, errflg=0;

        do {
                csr = inp(BA);                  /* read csr */
                mask = 0xc0;                    /* mask csr bits 5-0 */
                a = csr & mask;
                bog = *delay;
                if ((a == 0xc0) || (a == 0x40)) {   /* A/D ERR set? */
                        printf("******  A/D ERROR  ******\n\n");
                        exit(1);
                }
                if (a == 0x80) {                /* A/D DONE set? */
                        csr = inp(BA+2);        /* read low and high bytes */
                        a = inp(BA+3);
                        mask = (a << 8) + csr;
                        if (mask > sv) *delay = abs + ((mask - sv) * sf);
                        else {
                                if (mask < sv) *delay = abs - ((sv - mask) * sf);
                                else *delay = abs;
                        }
                        if (*delay > aaint) *delay = aaint;
                        if (*delay < 0) *delay = 0;
```

```
                        if ((bog != *delay) && (bog+1 != *delay) && (bog-1 !=
*delay)) {
                                resultarray[*ptr] = *delay;
                                (*ptr)++;
                        }
                        errflg++;
                }
        } while (errflg == 0);
}

printresults(array, ptr, fp, ab)
int array[], *ptr, ab;
FILE *fp;
{
        int i, x;

        fprintf(fp, "A-B = %d\n\n", ab);
        fprintf(fp, "New A-B Intervals -\n");
        x = (int) (((float)spitem[0].numframes * spitem[0].frameduration) *
1000.0);
        if (*ptr != 0) {
                for (i=0;i<(*ptr)-1;i++) fprintf(fp, "\t%d", (array[i] + x));
                fprintf(fp, "\n\nEnd Value = %d", array[(*ptr)-1] + x);
        } else fprintf(fp, "\tNo New Values.");
}

makerandomstart(aaint, abint, sf)
int aaint, *abint;
float sf;
{
        int q, flag;

        printf("Seeking out value. Please wait.....\n\n");
        do {
                fflush(stdout);
                flag = 0;
                *abint = rand()%aaint;
                if (*abint < (int)((spitem[0].frameduration *
(float)spitem[0].numframes) * 1000.0))
                        flag++; /* i.e. length of sound A */
                q = (aaint/2) + (int) (sf * 1000.0);
                if (*abint > q) flag++; /* i.e. within upper limit of nob */
                q = (aaint/2) - (int) (sf * 1000.0);
                if (*abint < q) flag++; /* i.e. within lower limit of nob */
        } while (flag != 0);
```

```
}

initlcd()
{
        int i;
        static int setup[6] = {0x38, 0x38, 0x38, 0x06, 0x0e, 0x01};

        for (i=0; i<6; i++) {
                outpw(port_c, 0x04);
                zdelay(10);
                outpw(port_b, setup[i]);
        }
        zdelay(300);
}

lcd(x)
char *x;
{
        int i, ch;

        outpw(port_c, 0x04); /* prepare lcd to receive text */
        zdelay(10);
        outpw(port_b, 0x01); /* clear screen */
        zdelay(10);
        outpw(port_c, 0x04); /* prepare lcd to receive text */
        zdelay(10);
        outpw(port_b, 0x80); /* address of start of first line */
        zdelay(10);
        if (strlen(x) > 25) {printf("string too long"); exit(1);}
        for (i=0; i<strlen(x); i++) {
                ch = *(x+i);
                outpw(port_c, 0x0c);
                zdelay(10);
                outpw(port_b, ch);
        }
}
```



Oscillogram and spectrogram of first "wa" stimulus from Experiment Six



Oscillogram of reference stimulus used throughout thesis (50ms signal correlated noise, ramped 10ms at onset and offset)

List of Tables

Page

3.1 *Summary of experimental manipulations* 63
4.1 *Mean intervals set by all subject groups* 79
4.2 *Transformed means for all subject groups* 80
5.1 *Mean produced inter-onset intervals across speakers* 101
5.2 *independent t-test results for interval differences* 103
5.3 *Mean intervals set in perception* 106
5.4 *Mean absolute deviations from isochrony* 110
5.5 *Production and perception intervals across speakers* 112
6.1 *Mean intervals set for "ones"* 121
6.2 *Mean intervals set for "twos"* 122
6.3 *Mean absolute deviations from isochrony - "ones"* 123
6.4 *Mean absolute deviations from isochrony - "twos"* 123
6.5 *P-centres for "ones" and "twos" stimuli* 126
6.6 *Vos and Rasch model predictions* 132
6.7 *Howell model predictions* 133
6.8 *Marcus model predictions* 135
6.9 *Speaker SR intervals* 143
6.10 *Speaker WC intervals* 143
6.11 *Speaker SM intervals* 143
6.12 *Speaker DG intervals* 144
6.13 *P-centres for "ones" and "twos" stimuli* 144
7.1 *Mean intervals set per stimuli pair* 158
7.2 *Mean absolute deviations from isochrony* 159
7.3 *Vos and Rasch model predictions* 163
7.4 *Howell model predictions* 165
7.5 *Marcus model predictions* 167
8.1 *Mean intervals for all stimuli* 178
8.2 *Mean absolute deviations from isochrony* 179

8.3 *P-centres of "eight" stimuli* 181
8.4 *Howell model predictions* 182
8.5 *Mean "eight" production intervals* 186
9.1a *Sonority values and examples* 195
9.1 *Mean "sha" intervals* 205
9.2 *Mean "wa" intervals* 205
9.3 *Mean "ae" intervals* 206
9.4 *Mean absolute deviations from isochrony for "sha" stimuli* 206
9.5 *Mean absolute deviations from isochrony for "wa" stimuli* 207
9.6 *Mean absolute deviations from isochrony for "ae" stimuli* 207
9.7 *Amount of ramping, rise time and P-centres for "sha" stimuli* 213
9.8 *Amount of ramping, rise time and P-centres for "wa" stimuli* 214
9.9 *Amount of ramping, rise time and P-centres for "ae" stimuli* 215
9.10 *Slope of regressions of P-centres against rise times* 217
9.11 *Marcus model predictions* 219
10.1 *Mean intervals set for onset ramped stimuli* 235
10.2 *Mean intervals set for offset ramped stimuli* 236
10.3 *Mean absolute deviations from isochrony - onset ramped stimuli* 237
10.4 *Mean absolute deviations from isochrony - offset ramped stimuli* 237
10.5 *P-centres of onset and offset ramped stimuli* 242
11.1 *Mean intervals set for duration varying stimuli* 256
11.2 *Absolute deviations from isochrony - duration varying stimuli* 257
11.3 *P-centres of duration varying stimuli* 259
11.4 *Marcus model predictions* 261
11.5 *Howell model predictions* 262
11.6 *Vos and Rasch model prediction* 263
12.1 *P-centres, FAIM predictions and Marcus model predictions* 289

List of Figures

Page

1.1 *Difference between perceptual and physical isochrony* 7
1.2 *Marcus P-centre model* 12
1.3 *The rhythm setting tasks of Marcus and Howell* 15
1.4 *Syllabic centre of gravity model of P-centre location* 17
2.1 *Vos and Rasch model of Perceptual Onset* 49
2.2 *Gordon model of Perceptual Attack Time* 54
4.1 *P-centre alignment task* 70
4.2 *Absolute deviations from isochrony - tapping and trial* 81
4.3 *Absolute deviations from isochrony - tempo and trial* 82
4.4 *Absolute deviations from isochrony for all four groups* 83
5.1 *Mean production inter-onset intervals across speakers* 102
5.2 *Oscillograms of "ones" and "twos" from different speakers* 105
5.3 *Mean intervals set using speech from different speakers* 108
5.4 *Perception ratios plotted against production ratios* 113
6.1 *Oscillograms of "ones" and "twos" from different speakers* 119
6.2 *P-centres of "ones" and "twos" against speaker* 126
6.3 *Production intervals from speaker PH* 128
6.4 *Vos and Rasch model predictions against P-centres* 132
6.5 *Howell model predictions against P-centres* 134
6.6 *Marcus model predictions against P-centres* 135
6.7 *Oscillograms of "ones" and "twos" from second group of speakers* 141
6.8 *P-centres of "ones" and "twos" from second group of speakers* 145
7.1 *Oscillograms of stimuli used by Tuller and Fowler 1981* 151
7.2 *Oscillograms of normal and infinitely peak clipped speech* 154
7.3.1 *Amplitude/time plot of output of ER-2 insert earphone* 156
7.3.2 *Amplitude/time plot of output of TDH-38 headphone* 156
7.4 *Mean absolute deviations from isochrony (IPC stimuli)* 159
7.5 *Vos and Rasch predictions against mean deviations* 164

7.6 *Howell model predictions against mean deviations* 166
7.7 *Marcus model predictions against mean deviations* 167
8.1 *Oscillograms of edited "eight" stimuli* 177
8.2 *Mean intervals set with "eight" stimuli* 180
8.3 *P-centres of "eight" stimuli* 181
8.4 *Mean "eight" production intervals for both speakers* 187
9.1 *Oscillograms of ramped "sha" stimuli* 198
9.2 *Oscillograms of ramped "wa" stimuli* 200
9.3 *Oscillograms of ramped "ae" stimuli* 202
9.4 *Oscillogram and spectrogram of first "sha" token* 203
9.5 *Subjects' mean deviations against "sha" stimuli combinations* 210
9.6 *Subjects' mean deviations against "wa" stimuli combinations* 210
9.7 *Subjects' mean deviations against "ae" stimuli combinations* 211
9.8 *"sha" P-centres against rise time* 214
9.9 *"wa" P-centres against rise time* 215
9.10 *"ae" P-centres against rise time* 216
9.11 *P-centres of "wa" and "ae", and "sha" against rise time* 217
9.12 *Marcus model predictions against all P-centres* 221
10.1 *Oscillogram and spectrogram of synthetic vowel* 231
10.2 *Oscillograms of onset ramped stimuli* 233
10.3 *Oscillograms of offset ramped stimuli* 233
10.4 *Standard deviations against means for both subjects' deviations* 240
10.5 *P-centres of onset and offset ramped stimuli* 244
11.2 *P-centres against stimulus duration* 260
11.3 *Howell, Marcus and Vos and Rasch predictions and P-centres against duration* 264
12.1 *P-centre modelling protocol* 274
12.2 *GammaTone filterbank output* 276
12.3 *Parameters for modelling* 278
12.4 *P-centre modelling*
12.5 *P-centre against frequency against rise time - 3-D plot* 282
12.6 *FAIM output* 284

12.7 *FAIM prediction against all thesis stimuli P-centres* 290

12.8 *Marcus model predictions against all thesis stimuli P-centres* 291