TR－H－062

THE PERCEPTION OF CONCURRENT VOWELS: PERIODIC AND APERIODIC VOWELS

Andrew P. Lea
Quentin Summerfield(MRC IHR)
Minoru Tsuzaki
Alain de Cheveigné(CNRS)

1994. 3. 7

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011
Facsimile: +81-7749-5-1008

THE PERCEPTION OF CONCURRENT VOWELS: PERIODIC AND APERIODIC VOWELS

Andrew P. Lea[1], Quentin Summerfield[2],

Minoru Tsuzaki[1] & Alain de Cheveigné[1&3]

[1]ATR Human Information Processing
Research Laboratories,
2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto 619-02,
Japan.

[2]MRC Institute of Hearing Research,
University Park,
Nottingham NG7 2RD,
United Kingdom.

[3]Laboratoire de Linguistic Formelle,
CNRS/Université Paris 7,
tc 806, 2 place Jussieu,
Paris 75251, France.

ABSTRACT

The first aim of the experiments described in this paper was to determine whether listeners can use the differences between periodically and aperiodically excited speech to segregate concurrent vowels which consist of one periodic and one aperiodic vowel. The second aim was to establish whether listeners can segregate either the periodic constituent, the aperiodic constituent, or both constituents of the concurrent vowels. In Experiment 1, 2, 3, 4 and 6 listeners identified both constituents of pairs of vowels presented concurrently. The constituents were either pulse-excited with fundamental frequencies (f0s) of 100 Hz and 112 Hz, or noise-excited. The results of Experiments 1, 2, 3, 4 and 6 show that listeners can use the difference in voicing between aperiodic and periodic constituents to segregate them. These experiments also show that the ability of listeners to use this difference in voicing is as good as a difference in pitch between two voiced vowels. Experiment 1 shows that listeners segregate the periodic/aperiodic concurrent vowel by only segregating only the aperiodic constituent and not the periodic constituent. However, Experiment 2 shows that both the periodic and aperiodic constituents were segregated. A number of procedural differences existed between Experiments 1 and 2 which might explain these contrasting results. Experiments 3 and 4 removed the possibility of all but one of these differences explaining the contrasting results of Experiments 1 and 2. The remaining difference was the method of matching the amplitudes of the periodic and aperiodic vowels. Therefore, Experiment 5 conducted a loudness matching experiment to find the most appropriate method of matching the periodic and aperiodic vowels. Experiment 6 used the results of Experiment 5 to show that when the vowels are correctly matched for equal loudness both the periodic and aperiodic vowels can be segregated. The implications for competing speech segregation are discussed.

I. INTRODUCTION

It is rare in the everyday world that we listen to speech without interfering sounds being present. Often the interference is the speech of another talker. For the purposes of this paper, the problem of separating speech from interfering sounds is simplified to separating two competing voices. Listeners are able to segregate a voice from interfering sounds by utilizing a number of different cues (Cherry, 1953; Darwin, 1981; Brokx *et al*, 1979; McAdams, 1988, Bregman, 1990). For example, a difference in fundamental frequency (f0) between two competing voices is common in the natural environment and its importance has been demonstrated as a cue for listeners. If two voices are presented simultaneously it is easier to understand what either talker is saying when the voices have different f0's (Brokx *et al*, 1979; Brokx and Nooteboom, 1982; Scheffers, 1983; Zwicker, 1984; Assmann and Summerfield, 1990; Chalikia and Bregman, 1989; Summerfield and Assmann, 1991; Culling and Darwin, 1993). A difference in f0 between two voices can be viewed as a difference between the periodic rates of excitation of the vocal chords during the production of each voice. Arguably, a more rudimentary difference is that between periodic speech (i.e. voiced speech) and aperiodic speech (i.e. whispered speech). This paper concentrates on the ability of listeners to use the difference between periodic and aperiodic speech to segregate concurrent voices.

A powerful and convenient paradigm with which to measure the ability of listeners to exploit cues for voice segregation was introduced by Scheffers (1983) and it is used in Experiments 1 to 4 described in this paper. His paradigm was designed to investigate listeners' ability to use a difference in pitch as a basis for segregation. He presented two synthetic vowels simultaneously and monaurally to listeners. Such pairs are referred to as "concurrent vowels" in this paper. Scheffers showed that listeners could identify both of the constituent vowels with an accuracy that was significantly above chance, even when there was no difference in f0 between them. In this condition listeners hear a voice producing a "dominant" vowel whose phonetic identity is colored by the impression of a second vowel. When a difference in f0 of about half a semitone (3%) or more is introduced, listeners often hear two voices producing vowels on different pitches. In addition identification accuracy improves by up to 20%, reaching a plateau for a difference of about 2 semitones (12%). It has been suggested that the components that define the formants of each vowel are grouped by their harmonicity or periodicity and assigned to the appropriate voices when the f0s are different (Broadbent and Ladefoged, 1957; Assmann and Summerfield, 1990). However, more recent results show that it is only f0 differences in the low frequency region that are useful for segregation (Culling and Darwin, 1993), suggesting that improvement in recognition is entirely due to listeners being able to segregate resolved harmonics which define the first formant (F1) and the second formant (F2) if it is low in frequency.

The first aim of the experiments described here was to determine whether listeners can exploit the difference in mode of excitation between periodic and aperiodic speech to segregate concurrent vowels. This is investigated in the perceptual experiments described in Sections II to VII below. The second aim was to establish how listeners segregated the concurrent vowels. This topic is discussed in Section VII.

II. EXPERIMENT 1

Experiment 1 matched the periodic and aperiodic vowels using excitation patterns (see below) to test the hypothesis that listeners can use the difference between periodic and aperiodic vowels to segregate them.

A. STIMULI OF EXPERIMENT 1

Stimuli were synthesized digitally (10000 sample/s; 16-bit amplitude quantitization). They were based on five single vowels which were steady-state exemplars of the British-English monophthongal vowels [a], [i], [ɜ], [u] and [ɔ] and therefore will be called the "English vowels". Their formant frequencies and bandwidths are listed in Table 1.

The stimuli for Experiment 1 were created by first creating an aperiodic (noise-excited) segment for each single vowel using the Klatt (1980) synthesizer in its cascade configuration. These segments were 1075-ms in duration. Each 1075 ms segment was divided into five 215 ms tokens. These tokens had slightly different amplitude spectra due to the random fluctuations in spectral amplitude of the noise source. Each token was processed in the same way as shown in Figure 1 using a token of the vowel [i]. Panel A of Figure 1 shows the waveform of a aperiodic [i] token produced directly by the Klatt synthesizer and Panel B shows the amplitude spectrum of the same token.

The aim of the processing was to produce periodic and aperiodic tokens of each vowel whose excitation patterns were matched as closely as possible at the frequencies of the harmonics of the periodic stimuli. It is not possible to achieve this goal using the Klatt synthesizer directly, because the spectra of the noise and pulse sources have different slopes (periodic: -6 dB octave; aperiodic: 0 dB/octave). Therefore, the following procedure was used. First, the waveforms of the aperiodic tokens were integrated to give a -6 dB/octave de-emphasis. This was necessary as without the -6 dB/octave de-emphasis the periodic vowels generated from the aperiodic vowels sounded unnaturally bright. The integrated waveform is shown in Panel C of Figure 1. Integration introduced some intense very low-frequency components which were not audible. The amplitude spectrum of the integrated waveform is shown in Panel D of Figure 1 and the excitation pattern (Moore and Glasberg, 1983, 1987) is plotted in Panel E. If Panels B and D are compared it can be seen that the higher frequencies have been attenuated compared to the lower frequencies. These tokens are the ones used in the experiment described below. Despite the intense low-frequency components the tokens sounded perfectly natural to listeners as the intense low-frequency components were not audible to listeners.

Formant	[a]	[i]	[ɜ]	[u]	[ɔ]
F1	658	269	425	281	362
F2	1001	2115	1440	1140	695
F3	2459	3130	2431	1992	2549
F4	3435	3517	3154	3047	3059
F5	3850	3850	3850	3850	3850

Table 1. Frequencies of the five formants used to synthesize the single-vowel constituents of the concurrent vowels. The 3-dB bandwidths (Hz) used were the default parameters of the Klatt (1980) synthesizer: 90, 110, 170, 250 and 300 for F1 through F5 respectively. These formant values were derived from the median values of four male speakers in a "now I'll say h/*vowel*/d again" context. The F5 shown is the default of the synthesizer.

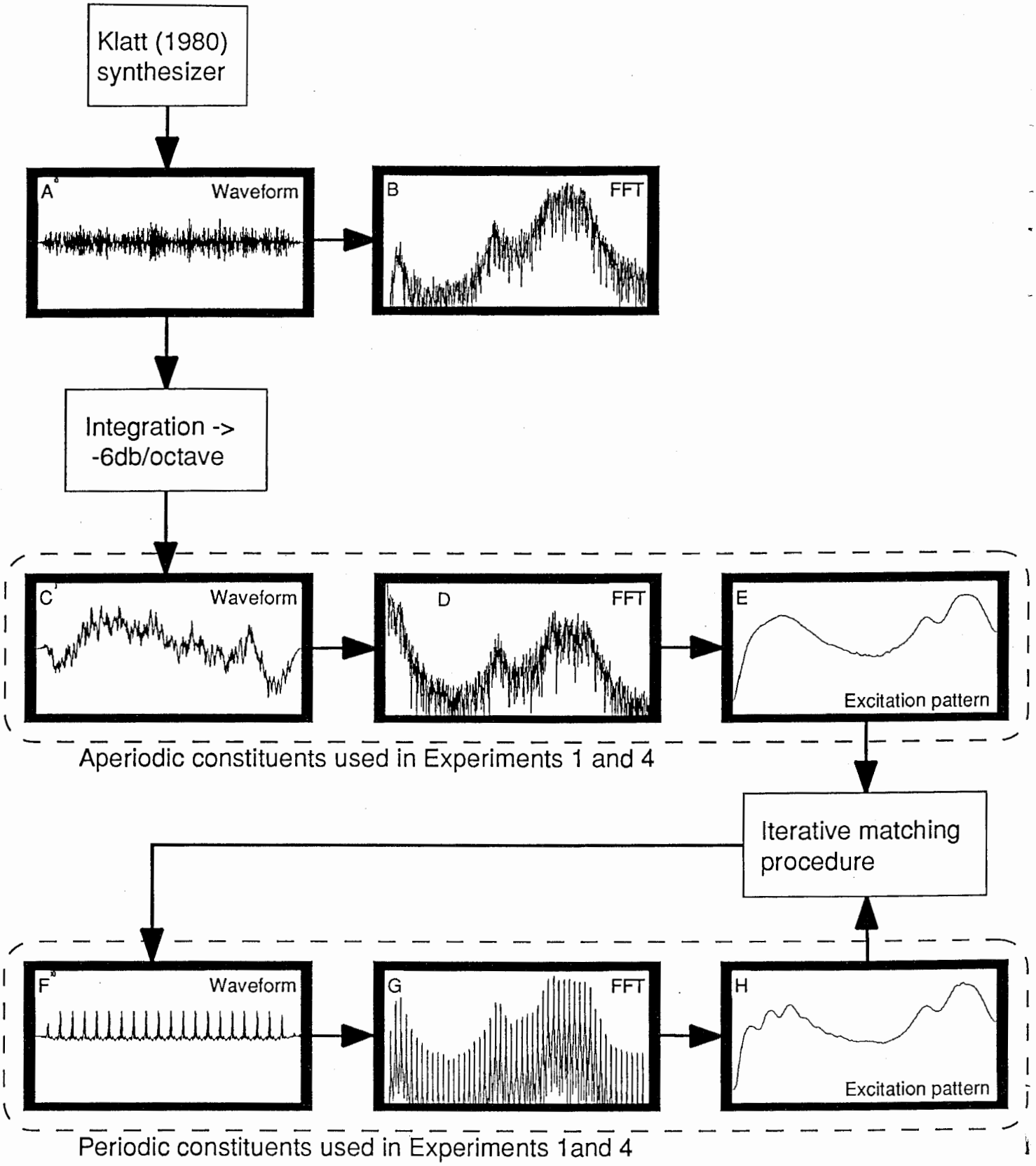

Figure 1. Creating the stimuli used in Experiment 1. Panel A shows the waveform of an aperiodic [i] token produced by the Klatt synthesizer. Panel B shows the amplitude spectrum of the token. The waveform in Panel A was integrated to produce the waveform in Panel C. Integration had the effect of reducing the spectral slope by 6 dB/octave to -6 dB/octave. The amplitude spectrum and excitation pattern of the integrated aperiodic vowel are shown in Panels D and E respectively. The excitation pattern of the integrated aperiodic token was used to derive a matching periodic vowel token using the iterative procedure described in Appendix A. The waveform, amplitude spectrum and excitation pattern of a periodic [i] token with a 100-Hz f0 which has been matched to the aperiodic [i] token are shown in Panels F, G and H respectively.

The excitation pattern of each aperiodic token was used to synthesize a periodic token with a matching excitation pattern, using an iterative matching procedure. The required f0, stimulus duration, phase spectrum and on-set and off-set window durations were included in the synthesis. The iterative matching procedure sought to minimize the differences between the excitation pattern of the aperiodic token as shown in Panel E and the excitation pattern of the periodic token as shown in Panel H. A full description of the iterative matching procedure can be found in Appendix A. A periodic token (Panel F) was derived from each of the five aperiodic tokens of each vowel, thus making five slightly different periodic and aperiodic tokens for each vowel. Examples of the results of this iterative matching procedure are shown in Figure 2 for the five tokens of the vowel [i] used in Experiment 1.

The rational for using excitation patterns as a basis for matching periodic and aperiodic tokens is as follows. An essential property of peripheral auditory analysis is to ensure that energy in different frequency bands generates neural excitation in different fibers of the auditory nerve. In this role as a frequency analyzer, it has been convenient to model the peripheral auditory system as an array of linear, overlapping, band-pass filters, with successive filters in the array tuned to different center frequencies (e.g. Helmholtz, 1863; Zwicker and Feldtkeller, 1967; Patterson, 1974; Moore and Glasberg, 1983). The frequency responses of these "auditory" filters have been estimated from the results of masking experiments (see Patterson and Moore, 1986, for a review) with the result that it is possible to program an array of such filters as a digital auditory filter bank (e.g. Assmann and Summerfield, 1990; Patterson and Holdsworth, 1990). The excitation pattern of a sound can be computed by presenting the waveform of the sound to each member of the array of filters and plotting the root mean squared levels of the waveforms emerging from the filters as a function of the center frequencies of the filters. In plotting excitation patterns, we have scaled the frequency axis in units of the equivalent rectangular bandwidths (erbs) of the filters. Equal increments along this scale correspond to equal distances of approximately 0.85 mm along the cochlear partition. For this reason, excitation patterns can be thought of as providing an estimate of the distribution of auditory excitation across place in the peripheral auditory system (Moore and Glasberg, 1986). Excitation patterns have proved to be more informative than Fourier spectra for understanding many aspects of the perception of complex sounds (e.g. Moore and Glasberg, 1986), including aspects of the perception of speech (e.g. Assmann and Summerfield, 1989). Therefore matching the periodic and aperiodic vowels using excitation patterns should provide the best match between them.

Periodic vowels were synthesized with two f0s: 100 Hz and nearly 2 semitones higher at 112 Hz. The phase spectrum produced by cascade synthesis (a phase shift across each formant frequency of 180^{o}) was used for each of the voiced vowels. The periodic vowels with a 100-Hz f0 are abbreviated to P_1, the 112-Hz voiced vowels are abbreviated to P_2 and the aperiodic vowels are abbreviated to A. All tokens were windowed using the

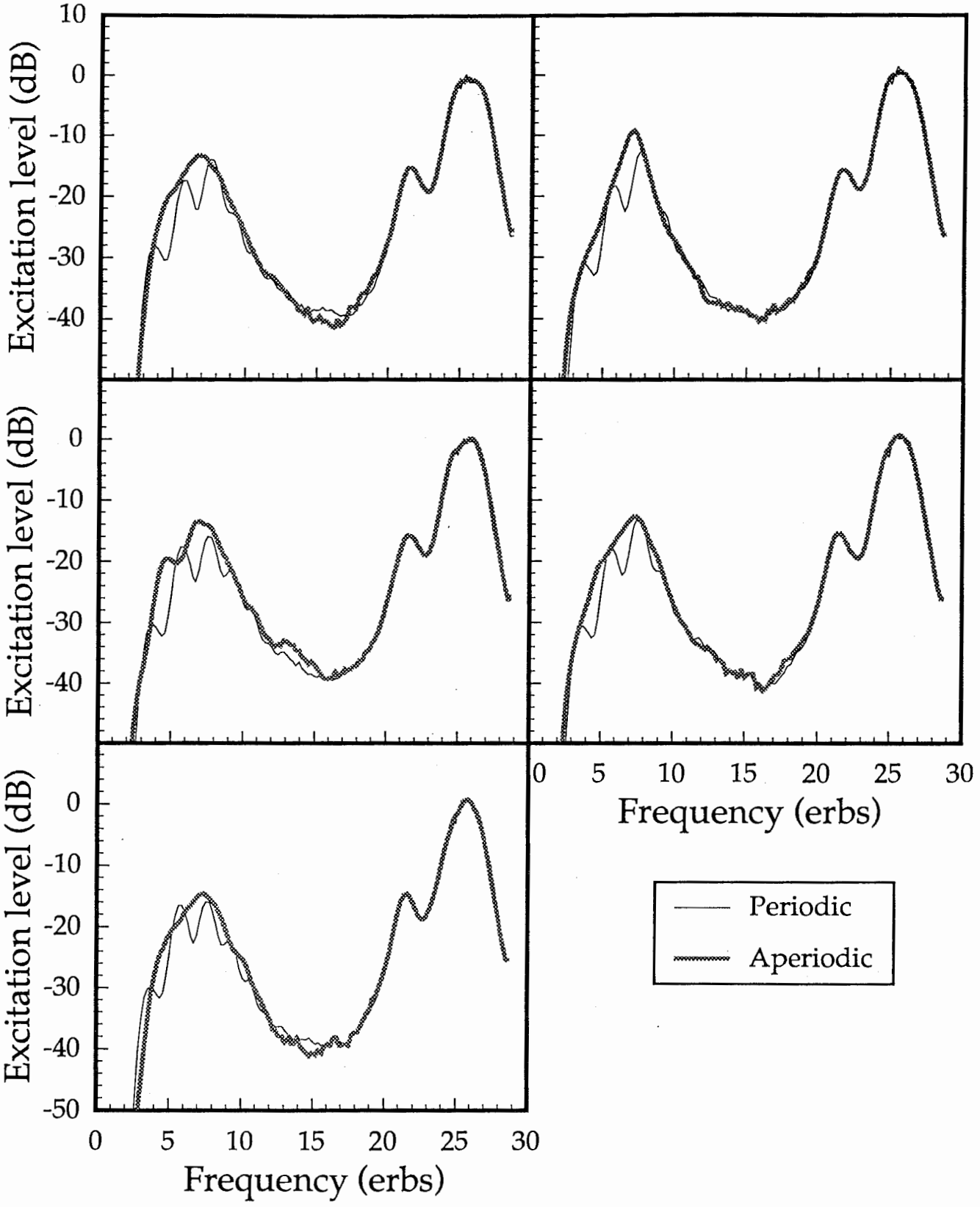

Figure 2. Excitation patterns of the five aperiodic [i] tokens and the five periodic [i] tokens with 100-Hz f0s used in Experiment 1. The standard deviations between the aperiodic and periodic tokens [i] token excitation patterns at the 200, 400, 800, 1600 and 3200 Hz points (5.37, 8.90, 13.59, 19.08 and 24.82 erbs) were 0.11, 0.81, 0.36, 0.05 and 0.30 dB respectively and the average difference (aperiodic minus periodic) at these points were 0.4, -0.2, 0.0, 0.0 and -0.1 dB respectively. The largest difference between any two matched tokens at any of these points was -0.9 dB.

two halves of a 20-ms raised-cosine function. The duration between the 6-dB-down points of the resulting waveforms was 195 ms.

The original versions of the processed aperiodic vowels were synthesized by the Klatt synthesizer with equal vocal intensity. This meant that there were differences in amplitude between the five vowels. The amplitudes of the voiced vowels were not changed from those produced by the iterative synthesis procedure described above. The rank order of the 100-Hz vowels averaged over the five tokens was [a], [i], [ɔ], [ɜ] and [u] with the last four vowels being 4.4, 4.5, 5.4 and 6.7 dB (RMS of the sample values) less intense than the average of the five tokens of [a] respectively.

Concurrent vowels were created by summing the corresponding samples of single vowels. Tokens of vowels were combined only with the corresponding tokens of other vowels; e.g., token three of the [i] vowel was only combined with other token threes. This ensured that concurrent vowels were fomed with vowel pairs which were derived from the same pseudo-random excitation spectrum. Thus five slightly different tokens of each concurrent vowel. Concurrent vowel tokens in which the same phonetic constituent vowel was present twice; "non-exclusive concurrent vowels", (Culling and Darwin, 1993) were not used in this study.

Four types of concurrent vowel were created. The first type consisted of two P_1 single-vowel constituents and is abbreviated to P_1P_1. The second type consisted of one P_1 constituent and one P_2 constituent and is abbreviated to P_1P_2. The third type consisted of one P_1 constituent and one A constituent and is abbreviated to P_1A. The final type consisted of two A constituents and is abbreviated to AA. An individual constituent of a concurrent vowel will identified as P_1/P_2 which means the P_1 constituent of the P_1P_2 concurrent vowel and P_2/P_1 means the P_2 constituent of the P_1P_2 concurrent vowel.

B. PROCEDURE OF EXPERIMENT 1

The stimuli were presented on-line (Masscomp 5600), low-pass filtered at 4.5 kHz (KEMO VBF/8, -135 dB/octave), and presented to the left ear phone of a Sennheiser HD-414x headset.

The average presentation level of the single-vowel stimuli was 72 dB(A) for the periodic vowels and 73 dB(A) for the aperiodic vowels. The average presentation level of the concurrent-vowel stimuli was 78 dB(A).

Listeners were tested individually in a sound attenuated room. They were asked to give two different responses to each of the concurrent-vowel stimuli, and responded by pressing VDU keys labeled with the orthographic representations of the five single vowels. Feedback was presented on the VDU screen to indicate the correct response, after the listener had responded to each stimulus, using the same orthographic representation. Listeners received experience of the experimental stimuli by responding to a sequence of trials in which a token of each stimulus appeared once. In the experiments, each concurrent-vowel token was presented twice, making a total of ten presentations of each concurrent vowel. Two regimes of presentation were used, one where the stimulus on

each trial was presented only once to the listeners, called "single-shot" presentation and the second where listeners were allowed as many presentations of each stimulus as they wanted before responding, called "multi-shot" presentation. Two experimental sessions were used to present stimuli for both of the two presentation paradigms. The stimulus presentation order was randomized across the four different types of stimuli for each presentation paradigm. Six listeners were tested with the single-shot paradigm first followed by the multi-shot paradigm. The other three were tested in the reverse order.

The nine listeners were native British-English speakers and included the first two authors. They were either staff or students at the University of Notthingam and were paid for their participation. All were adults with pure-tone audiometric thresholds within 15 dB of the ANSI standard in each ear (ANSI, 1969).

C. RESULTS OF EXPERIMENT 1

The listeners were first tested with the single vowels in isolation to establish that they could identify them accurately. The identification level, averaged over the nine listeners, was 99.0% for the P_1 vowels, 99.1% for the P_2 vowels and 98.9% for the A vowels.

The results for both presentation regimes of the main experiment were scored in two ways. The first used the "combinations-correct score" which has been used previously (e.g., Scheffers, 1983; Assmann and Summerfield, 1990). The combinations-correct score was computed by calculating the percentage of trials on which listeners correctly identified both constituents of the concurrent-vowel stimuli correct. The second method used the "constituents-correct score", which was used previously by Scheffers (1983), was computed by calculating the percentage of trials on which each constituent of the concurrent-vowel stimuli was correctly recognized. Thus, if the stimulus was a combination of an [i] and an [a] and the response was 'AH+EE", the combinations-correct score was incremented by one, as were the constituents-correct scores for both the [a] and the [i]. If, on the other-hand, the response was 'AH+ER', only the constituents-correct score for the [a] was incremented. Thus, the constituents-correct score can tell us how well a constituent is segregated from the mixture according to its structure or the structure of the constituent that accompanies it.

A repeated measures analysis of variance (ANOVA) was performed with order of presentation regime (single-shot before multi-shot, or *vice versa*) as a between-listeners factor, and the within-listeners factors of regime (single-shot or multi-shot) and combination identification accuracy (P_1P_1, P_1P_2, P_1A and AA). Order of presentation regime was not significant ($F_{1,7}=3.1$), nor were any significant differences found between the single and multi-shot regimes ($F_{1,7}=0.0$), however, there proved to be a significant interaction between order of regime presentation and single-shot versus multi-shot regimes ($F_{1,7}=19.0$, $p<0.003$) and all other interactions were not significant. This outcome suggests that listeners learnt from which ever regime was presented first and improved their identification scores for the second regime presented.

Since, there were no interesting significant differences between the two presentation regimes, the results from both conditions are pooled in Figure 3 which shows the results averaged over listeners and concurrent vowels. In accordance with the principles of repeated measures analyses (e.g. Winer *et al* , 1991), where effects of treatments for a subject are measured relative to the average score produced by that subject across all treatments, the error bars in all Figure 3 and all other results figures of this paper, plot +/- one "intra-listener" standard deviation; i.e. the standard deviation of thresholds computed after subtracting the mean threshold of each subject from his/her individual thresholds.

A second repeated measures ANOVA using the pooled data was computed, it had the factor of condition. The ANOVA showed that the four conditions did differ significantly from their mean ($F_{3,24}=7.9$, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some concurrent vowel combinations differed from other concurrent vowel combinations. Accuracy of identification of the P_1P_2 concurrent vowel was higher than for the P_1P_1 concurrent vowel ($p<0.05$) and accuracy of identification of the P_1A concurrent vowel was higher than the P_1P_1 concurrent vowels ($p<0.01$). Accuracy of identification of the P_1P_1 and AA concurrent vowels did not differ. Accuracy of identification of the P_1P_2 and P_1A concurrent vowels did not differ.

The results were pooled over the single-shot and multi-shot regimes as no interesting differences were found for the combinations-correct score and are shown in Figure 4. A repeated measures ANOVA using the pooled data was computed, it had the factor of condition. The ANOVA showed that the six constituent types did differ significantly from their mean ($F_{5,40}=9.9$, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some constituent types differed from other constituent types. Accuracy of identification for the P_1/P_2 constituent type was higher than for the P_1/P_1 constituent type ($p<0.05$). Accuracy of identification for the A/P_1 constituent type was higher than for the A/A constituent type ($p<0.01$). Accuracy of identification for the P_1/A constituent type did not differ from the P_1/P_1 constituent type. Accuracy of identification for the A/P_1 constituent type was higher than for the P_1/A constituent type ($p<0.05$).

D. DISCUSSION OF EXPERIMENT 1

Before we enter into a discussion of the results of Experiment 1, it is necessary to define a few important terms using the terminology of Bregman (1990) for processes that the auditory system performs. Bregman (1990) calls the problem of separating the many sounds that we hear simultaneously "auditory scene analysis". The aim of auditory scene analysis is to separate the internal spectrogram so that separate physical events can be labeled individually. One level of representation in this process is the "stream". "An auditory stream is our perceptual grouping of the parts of the neural spectrogram which go together" (Bregman, 1990, page 9). Therefore a stream is the perceptual unit which refers to a single happening in the external physical world. In the process of stream formation, Bregman (1990) makes a distinction between the use of perceptual mechanisms which are hard-wired and those which are learnt. The use of hard-wired perceptual

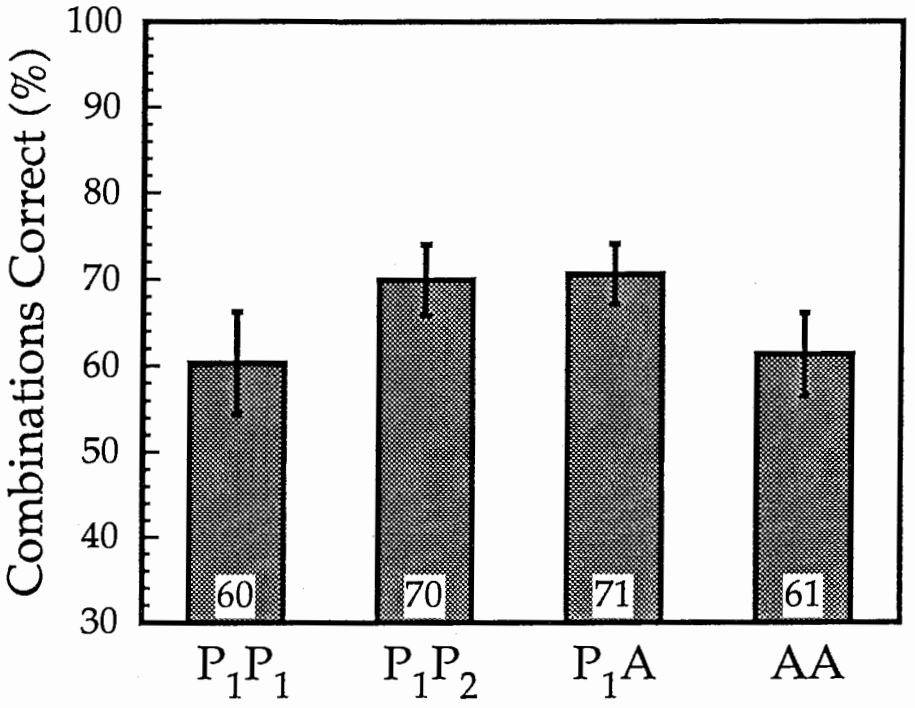

Figure 3. The results from Experiment 1 scored using the combinations-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results are averaged over the nine listeners and the different concurrent vowel combinations. In Panel A the results from the single-shot regime are plotted by closed symbols and the results from the multi-shot regime are plotted by the open symbols. In Panel B the results have been averaged over both presentation regimes.

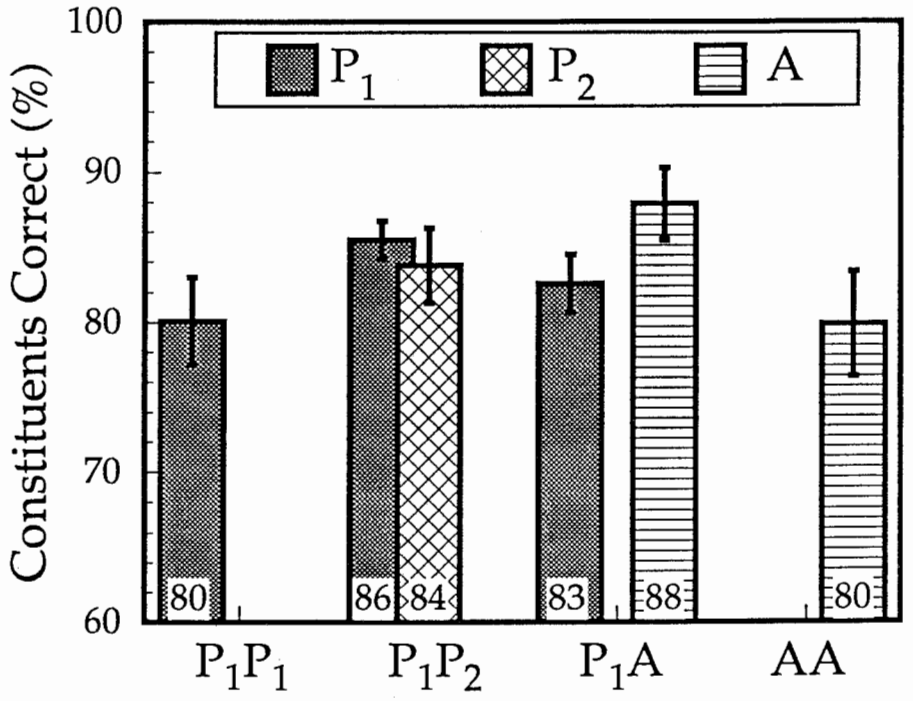

Figure 4. The results from Experiment 1 scored using the constituents-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the nine listeners and the different concurrent vowel combinations. In Panel A the results from the single-shot regime are plotted by closed symbols and the results from the multi-shot regime are plotted by the open symbols. In Panel B the results have been averaged over both presentation regimes

mechanisms he calls "primitive segregation", whereas the use of learned perceptual mechanisms he calls "schema-based segregation".

Accordingly, segregation will refer to the low-level primitive, probably innate, process of separating sounds according to their physical properties into different streams (Bregman, 1990). Whereas vowel identification will refer to the use of learned schema to recognize vowels (Bregman, 1990). Thus, it is assumed that segregation and identification are two separate, but not necessarily independent processes. Also, it is assumed that the process of identification relies on the products of segregation having occurred.

For both the P_1P_1 and AA concurrent vowels there is no cue for segregation and since listeners' identification levels are much greater than chance level this must be the product of the schema-driven process of vowel identification. If listeners' identification level for the P_1P_2 and P_1A concurrent vowels are greater than those found for the P_1P_1 and AA concurrent vowels, we can infer that the primitive process of segregation has occurred. Hence increases in listeners' identification levels above the baselines of the P_1P_1 and AA concurrent vowels can be taken as an indication that the primitive process of segregation has occurred.

Since the accuracy of identification of the P_1P_2 combination was higher than for the P_1P_1, this shows that listeners were able to use the difference in pitch to segregate the constituents, thus replicating the finding that a difference in f0 between concurrent vowels leads to increased accuracy of identification e.g. Scheffers (1983). Also, the accuracy of identification accuracy of the P_1A concurrent vowel was higher than for the P_1P_1 and AA concurrent vowels, this shows that listeners were able to use the difference in voicing to segregate the constituents. The increase in identification accuracy for the P_1P_2 and the P_1A concurrent vowels over the P_1P_1 and the AA concurrent vowels was of a similar size. This result suggests that listeners are able to use a difference in excitation between concurrent vowels as effectively as a difference in pitch between competing voices to segregate the concurrent vowels. Also, listeners can identify the AA concurrent vowels as accurately as the P_1P_1 concurrent vowels thus suggesting that both A and P_1 vowels are equally identifiable for the listeners. Scheffers (1983) also used concurrent vowels in which both constituents were aperiodic. The relationship between the results of Scheffers and the results of the experiments described here is covered in the General Discussion section.

The identification accuracy of the P_1/P_2 constituent type was higher than that of the P_1/P_1 constituent type, but the identification accuracy of the P_2/P_1 constituent type was not higher than that of the P_1/P_1. It is probably the case that both constituents were segregated from each other, but the experiment did not have the statistical power to show a significant difference between the identification accuracy of the P_1/P_1 and P_2/P_1 constituent types.

The accuracy of identification of the A/P_1 constituent was higher than that of the A/A constituent, suggesting that the A/P_1 constituent was segregated from the P_1/A constituent. However, the accuracy of identification of the P_1/A constituent was no

different to that of P_1/P_1 constituent suggesting that the P_1/A constituent was not segregated from the A/P_1 constituent.

Before we use the results of Experiment 1 to try and infer the segregational strategy that listeners are using (see the General Discussion below) we must consider how the periodic and aperiodic constituents were matched in amplitude. Although matching using the excitation patterns of the vowels ought to provide the most accurate match possible, there is a possible problem with the iterative matching technique described in Appendix A. In the excitation patterns of the vowels the harmonics of the periodic vowels were matched to be the same amplitude as the continuous spectra of the aperiodic vowels. For the periodic vowels there is no energy between the harmonics. Therefore, the aperiodic vowels are more intense than the periodic vowels where the harmonics of the periodic vowels are resolved by the excitation pattern representation, this can be seen in Figure 2. In fact the aperiodic vowels are on average 9.6 dB more intense than the periodic vowels measured by the RMS (root of the mean squared) of the sample values. Thus, in low-frequency regions the aperiodic vowels are more intense than the periodic vowels. This is the very region in which Culling and Darwin (1993) have suggested that a difference in f0 between two periodic vowels causes segregation to occur.

The method of matching the periodic and aperiodic vowels may have produced a mismatch in amplitude which would mean that the aperiodic vowels would mask the periodic vowels more than *vice versa*. in the P_1A concurrent vowels. Thus, the result that the A/P_1 constituent was segregated and the P_1/A constituent was not segregated might be explained by the difference in masking. So, we cannot accept the results of Experiment 1 and use them to infer the strategy that listeners used to segregate the concurrent vowels.

A better method of matching the excitation patterns of the periodic and aperiodic vowels may have been to ensure that the RMS difference between the excitation patterns at not just harmonic frequencies, but also at frequencies between harmonics, was minimized. This would have increased the amplitude of the resolved harmonics of the periodic vowel, thus equalizing the levels of the periodic and aperiodic vowels in the low frequency region.

Clearly the method of matching the amplitudes periodic and aperiodic vowels is important and therefore another method which matched the vowels by the RMS dB of their sample values is used in Experiment 2.

III. EXPERIMENT 2

Experiment 2 was a replication of Experiment 1 except for a number of procedural differences. The main differences were that a different method of matching the periodic and aperiodic vowels was used; Japanese vowels and listeners were used and feedback was not used although it was used in Experiment 1.

A. STIMULI OF EXPERIMENT 2

Stimuli were synthesized digitally (10000 sample/s and then up sampled to 44100 sample/s; 16-bit amplitude quantitization). They were based on the five single vowel Japanese vowels [a], [i], [e], [u] and [o] and therefore will be called the "Japanese vowels". Their formant frequencies are shown in Table 2. The same bandwidths that were used in Experiment 1 were used again here.

Both the periodic and the aperiodic vowels were created using a cascade formant synthesizer which was based on the Klatt's (1980) synthesizer. Each vowel was created with three types of excitation. Aperiodic excitation was used to create aperiodic vowels, which are again abbreviated as A. Periodic excitation was used to create periodic vowels with a 99-Hz f0, which are abbreviated as P_1. Finally, periodic excitation was again used to create periodic vowels with 112-Hz f0, which are abbreviated to P_2. The duration of the vowels was 220 ms including 20 ms cosine shaped rise-fall windows giving the same duration as in Experiment 1. The vowels were matched so that the RMS of the sample values of each vowel were equal. Figure 5 shows the excitation patterns of a periodic [a] and an aperiodic [a] for comparison.

The same four types of concurrent vowel were created as with Experiment 1: P_1P_1, P_1P_2, P_1A and AA. Again only exclusive concurrent vowels were used. The constituents were added together so that both constituents were equal in intensity.

B. PROCEDURE OF EXPERIMENT 2

The stimuli were presented on-line (Macintosh IIfx controlling a MIDI-sampler) and presented to either the left or both ear phones of a STAX SR-Λ headset. The average presentation level for the constituents of the single vowels was 73, 73 and 74 dB(A) for the P_1,P_2 and A vowels respectively.

All seven listeners were Japanese female adults with normal hearing and were experienced with listening experiments. Training with feedback was given to each listener for the individual vowels until the hundred percent level was attained. No feedback was given during the concurrent vowel experimental sessions. Listeners were asked to give two different responses to each of the concurrent vowels. Each concurrent vowel was presented five times in an experimental session and two sessions of an hour each were required to gather the data.

C. RESULTS OF EXPERIMENT 2

Again the results were scored using both the combinations-correct and constituents-correct scores (see the results section of Experiment 1). The results of the combinations-correct score were tested to see if the two different types of presentation of diotic and monotic were significant. An ANOVA was performed with the factor of diotic vs. monotic presentation, it was not significant ($F_{1,47}$=0.01). Therefore, the results were summed over both types of presentation.

Formant	[a]	[i]	[e]	[u]	[o]
F1	650	250	450	250	550
F2	1050	2250	1950	850	850
F3	2850	3050	2650	2050	2250

Table 2. Frequencies of the three formants used to synthesize the single-vowel constituents of the concurrent vowels. The 3-dB bandwidths of the formants were 90, 110 and 170 Hz for F1 through to F3 respectively. These formant values were chosen so that the vowels sounded as natural as possible to Japanese listeners.

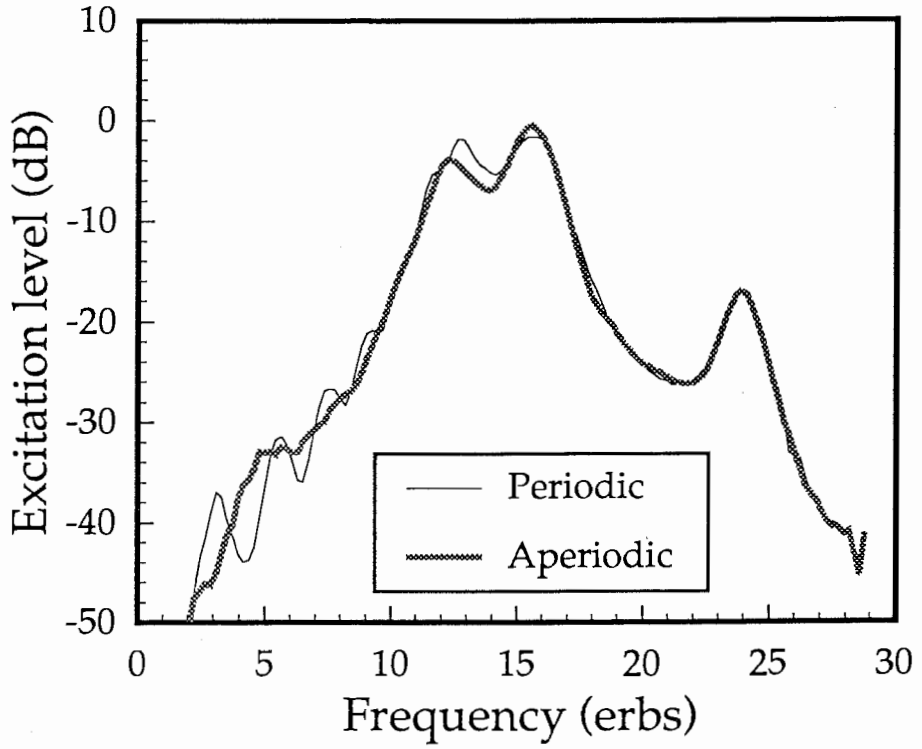

Figure 5. Excitation patterns of the vowel [a] used in Experiment 2. The black line shows the excitation pattern of the periodic vowel with a 99-Hz f0. The grey line shows the excitation pattern of the aperiodic vowel. The two vowels were matched in amplitude by making the RMS of the sample values of each vowel the same.

Figure 6 shows the results scored using the combinations-correct score averaged over the seven listeners. A repeated measures ANOVA was performed with the factor of concurrent vowel type. Concurrent vowel type was found to be significant ($F_{3,18}$=22.3, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some concurrent vowel combinations differed from other concurrent vowel combinations. Accuracy of identification of the P_1P_2 concurrent vowel was higher than for the P_1P_1 condition ($p<0.01$). Accuracy of identification of the P_1A condition was higher than the P_1P_1 and AA conditions (both $p<0.01$). Accuracy of identification of the P_1P_1 concurrent vowels was higher than for the AA concurrent vowels ($p<0.01$).

Figure 7 shows the results scored using the constituents-correct score averaged over the seven listeners. The results have again been averaged over both diotic and monotic presentation types. A repeated measures ANOVA was performed with the factor of constituent type. Constituent type was found to be significant ($F_{5,30}$=22.2, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some constituent types differed from other constituent types. Accuracy of identification for the P_2/P_1 constituent type was higher than for the P_1/P_1 constituent type ($p<0.01$). Accuracy of identification for the A/P_1 constituent type was higher than for the A/A constituent type ($p<0.01$). Accuracy of identification for the P_1/A constituent type was higher than for the P_1/P_1 constituent type ($p<0.01$).

D. DISCUSSION OF EXPERIMENT 2

The results scored using the combinations-correct score show that listeners were able to use the excitation difference between periodic and aperiodic vowels to segregate both of them. The evidence for this can be seen in Fig. 7 where the P_1A is identified more accurately than the AA and the P_1P_1 concurrent vowels. The increase in identification accuracy for the P_1A concurrent vowels over the P_1P_1 and the AA concurrent vowels is of a similar order of magnitude to the increase in identification accuracy for the P_1P_2 concurrent vowels over the P_1P_1 concurrent vowels. These results are similar to those of Experiment 1. It should be noted that the AA concurrent vowels are not identified as accurately as the P_1P_1 concurrent vowels, which is a different pattern to that shown in Experiment 1. This difference will be discussed in the General Discussion.

The results scored using the constituents-correct score show that listeners were able to segregate both the voiced vowel and the whispered vowel from the P_1A concurrent vowel. The evidence for this is shown in Figure 7 where the P_1/A is identified more accurately than the P_1/P_1 and the A/P_1 is identified more accurately than the A/A.

This last result is different to that found in Experiment 1 where listeners could only segregate the A/P_1 and not the P_1/A constituents. There are a number of procedural differences between Experiments 1 and 2, besides the difference in matching methods, which might possibly have produced this difference in the results. Firstly, in Experiment 1 five slightly different tokens of each constituent were used. Secondly, feedback was given in Experiment 1, but not in Experiment 2, so listeners might have learnt as the experiment

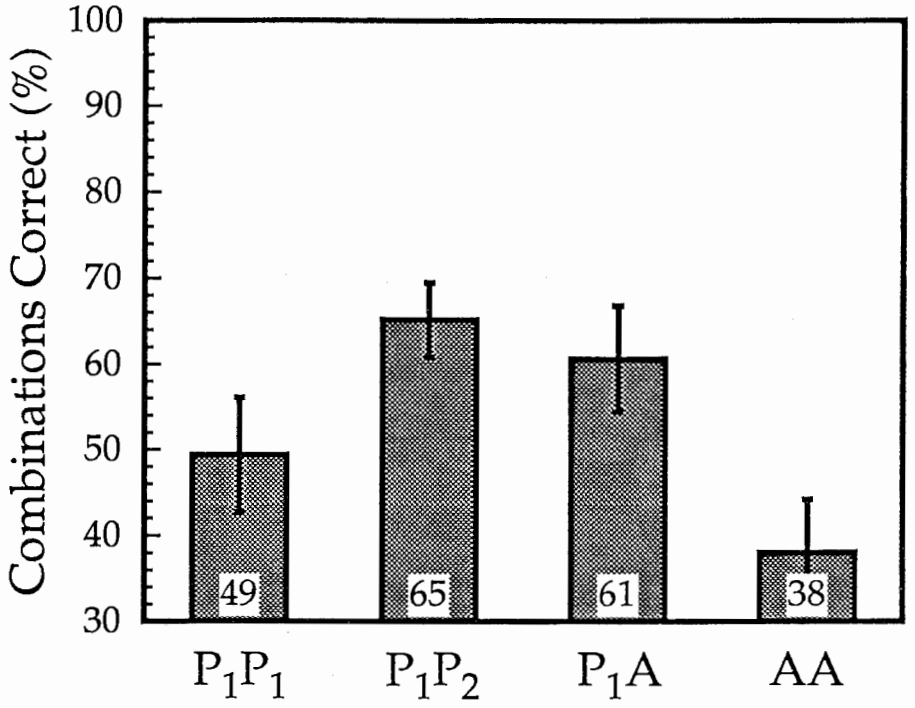

Figure 6. The results form Experiment 2 scored using the combinations-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

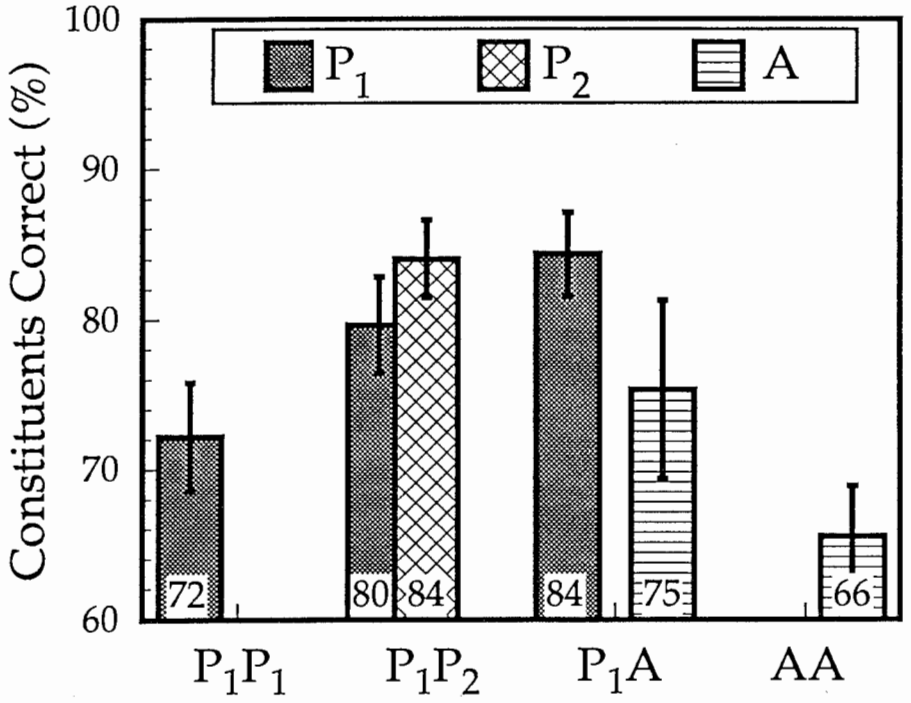

Figure 7. The results from Experiment 2 scored using the constituents-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

progressed. Thirdly, all the constituents in Experiment 2 had the same amplitude in terms of dB RMS, but in Experiment 1 the relative differences in amplitude between the constituents of a concurrent vowel were kept as produced by the Klatt synthesizer, thus assuring a difference in amplitude between the constituents of all the concurrent vowels. Fourthly, the formant frequencies of the English and Japanese vowel sets are different which might have produced different patterns of masking.

The effect of a difference in f0 between concurrent vowels has been studied in a number of different languages – Dutch (Scheffers, 1983), English (Assmann and Summerfield, 1991; Culling and Darwin, 1993), Japanese (current paper – Experiments 2 & 3) and French (de Cheveigné, personal communication). The different languages and hence different formant frequencies have affected the results in no discernible way. Therefore, it is unlikely that the different formant frequencies used between Experiments 1 and 2 could have produced this difference in results.

The possible affect of amplitude differences between the constituents of the concurrent vowels is examined in Experiment 3. Experiment 4 examines the possible effects of feedback and multiple tokens of each vowel. Experiment 5 is a loudness matching experiment to examine which method of matching the amplitude of the voiced and whispered should produce the most accurate results.

IV. EXPERIMENT 3

Experiment 3 uses the Japanese vowel to examine the effects of having amplitude differences between the constituents of concurrent vowels. In Experiment 1 amplitude differences between the five vowels were produced by the Klatt synthesizer. These amplitude differences meant that for all the concurrent vowels the constituents differed in amplitude. In Experiment 2 all five vowels were equalized in terms of RMS energy and therefore in the concurrent vowels all constituents had the same amplitude. To ensure that this difference did not produce the difference in results between the experiments, Experiment 3 was the same as Experiment 2, but introduced similar amplitude differences to those found in Experiment 1.

A. STIMULI AND PROCEDURE OF EXPERIMENT 3

The same stimuli and procedures as used in Experiment 2 were used here, except that amplitude differences were introduced between the five constituents of the concurrent vowels. The rank order of the amplitude of the vowels was [i], [a], [e], [o] and finally [u] with the vowels being 2.0, 4.0, 6.0 and 8.0 dB less intense than the [i] vowel respectively. This is a slightly different order than found in Experiment 1 and also the range of intensities is larger here.

The listeners used in Experiment 3 were the same as those used in Experiment 2.

B. RESULTS OF EXPERIMENT 3

Again the results were scored using both the combinations-correct and constituents-correct scores (see the results section of Experiment 1). The results of the combinations-

correct score were tested to see if the two different types of presentation of diotic and monotic were significant. An ANOVA was performed with the factor of diotic vs. monotic presentation, again it was not significant ($F_{1,55}$=1.1). Therefore, the results were summed over both types of presentation.

Figure 8 shows the results scored using the combinations-correct score averaged over the seven listeners. A repeated measures ANOVA was performed with the factor of concurrent vowel type. Concurrent vowel type was found to be significant ($F_{3,18}$=41.5, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some concurrent vowel combinations differed from other concurrent vowel combinations. Accuracy of identification of the P_1P_2 concurrent vowel was higher than for the P_1P_1 condition ($p<0.01$). Accuracy of identification of the P_1A concurrent vowel was higher than the P_1P_1 and AA concurrent vowels ($p<0.05$ and $p<0.01$ respectively). Accuracy of identification of the P_1P_1 concurrent vowel was higher than for the AA concurrent vowel ($p<0.01$). Accuracy of identification of the P_1P_2 and P_1A concurrent vowels did not differ.

Figure 9 shows the results scored using the constituents-correct score averaged over the seven listeners. The results have again been averaged over both diotic and monotic presentation types. A repeated measures ANOVA was performed with the factor of constituent type. Constituent type was found to be significant ($F_{5,30}$=24.4, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some constituent types differed from other constituent types. Accuracy of identification for the P_2/P_1 constituent type was higher than for the P_1/P_1 constituent type ($p<0.01$). Accuracy of identification for the A/P_1 and A/A constituent types did not differ. Accuracy of identification for the P_1/A constituent type was higher than for the P_1/P_1 constituent type ($p<0.01$).

To see if their were are difference in the results between Experiment 2 and 3 two further ANOVA were performed one for the combinations-correct score and one for the constituents-correct score. For the combinations-correct score the factors of experiment, and concurrent vowel type were used. Experiment was not significant ($F_{1,6}$=0.8) and experiment interacting with concurrent vowel type was not significant ($F_{3,18}$=0.6). Only tests with the factor of experiment are reported as the other factor is covered in the ANOVA for the combinations-correct score.

For the constituents-correct score factors of experiment and constituent type were used. Experiment was not significant ($F_{1,6}$=1.1) and experiment interacting with constituent type was significant ($F_{5,34}$=3.6, $p<0.01$). Again only tests with the factor of experiment are reported as the other factor is covered in the ANOVA for the constituents-correct score.

C. DISCUSSION OF EXPERIMENT 3

If we just examine the results of Experiment 3, it can be seen that the pattern is similar to Experiment 2. Figure 8 score shows that both the P_1P_2 and the P_1A concurrent-vowel types are identified more accurately than the P_1P_1 and AA concurrent-vowel types,

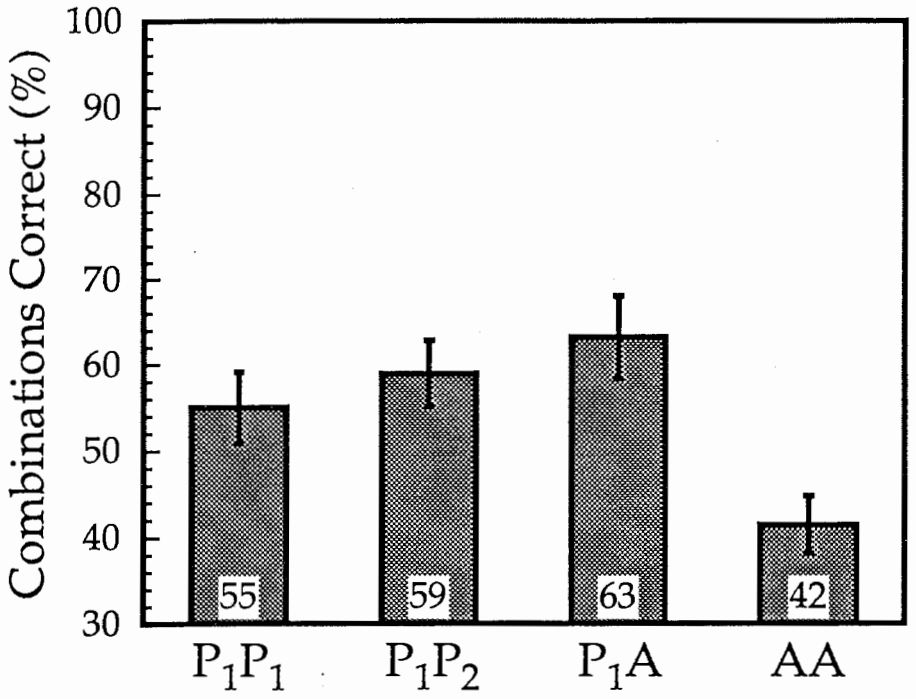

Figure 8. The results form Experiment 3 scored using the combinations-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

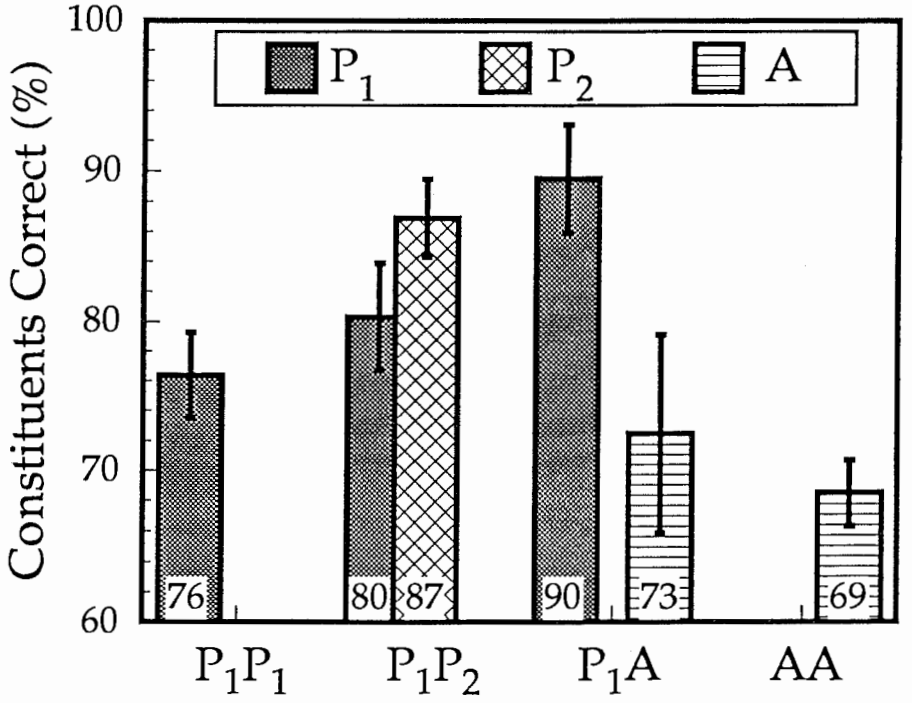

Figure 9. The results from Experiment 3 scored using the constituents-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

thus showing segregation has occurred. The pattern of results for the constituent correct score is not the same as for Experiment 2 also. Only the P_1/A constituent is higher than the P_1/P_1 constituent, however, the A/P_1 constituent is not higher than the A/A constituent showing that only the P_1/A has been segregated.

The ANOVA which compares the constituents-correct score results of Experiments 2 and 3 shows an interactions between constituent type and experiment. This interaction suggests that the amplitude differences between the vowels did affect the ability of listeners to segregate the constituents. However, although the amplitude differences between the constituents of Experiment 3 do produce noticeable effects, they cannot account for the difference in results between Experiments 1 and 2. The reason for this is that the difference in amplitudes only affected the segregation of the A/P_1 constituent and not the P_1/A constituent, it was the segregation of this constituent that differed between Experiments 1 and 2. Therefore a different cause must be searched for which explains the difference in results between Experiments 1 and 2.

V. EXPERIMENT 4

Experiment 4 uses the English vowels to ensure that the lack of feedback and using only single tokens of each concurrent vowel did not produce the difference in results between Experiments 1 and 2.

A. STIMULI AND PROCEDURE OF EXPERIMENT 4

The same stimuli as used in Experiment 1 were used here, the only difference being that only a single token of each concurrent vowel was used.

The stimuli were presented on-line (Macintosh IIfx controlling a MIDI-sampler) and presented to both ear phones of a STAX SR-Λ headset. The presentation level of the P_1 vowels was the same as in Experiment 2 72.8 dB(A). The other vowels and concurrent vowels were kept at the same relative intensities to the P_1 vowels as in Experiment 1.

The vowels were presented diotically. A different set of six listeners were used, who were all mother-tongue speakers of English and all had normal hearing.

Again the listeners were trained on the single vowels with feedback until a hundred percent accuracy was achieved. Each concurrent vowel was presented five times in each experimental session and two sessions were required to gather the data. No feedback for the concurrent vowels was given to listeners.

B. RESULTS OF EXPERIMENT 4

Again the results were scored using both the combinations-correct and constituents-correct scores (see the results section of Experiment 1). The results of the combinations-correct score are presented in Figure 10 averaged over the six listeners. A repeated measures ANOVA was performed with the factor of concurrent vowel type. Concurrent-vowel type was found to be significant ($F_{3,15}=17.9$, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some concurrent vowel combinations differed from other concurrent vowel combinations. Accuracy of identification of the P_1P_2

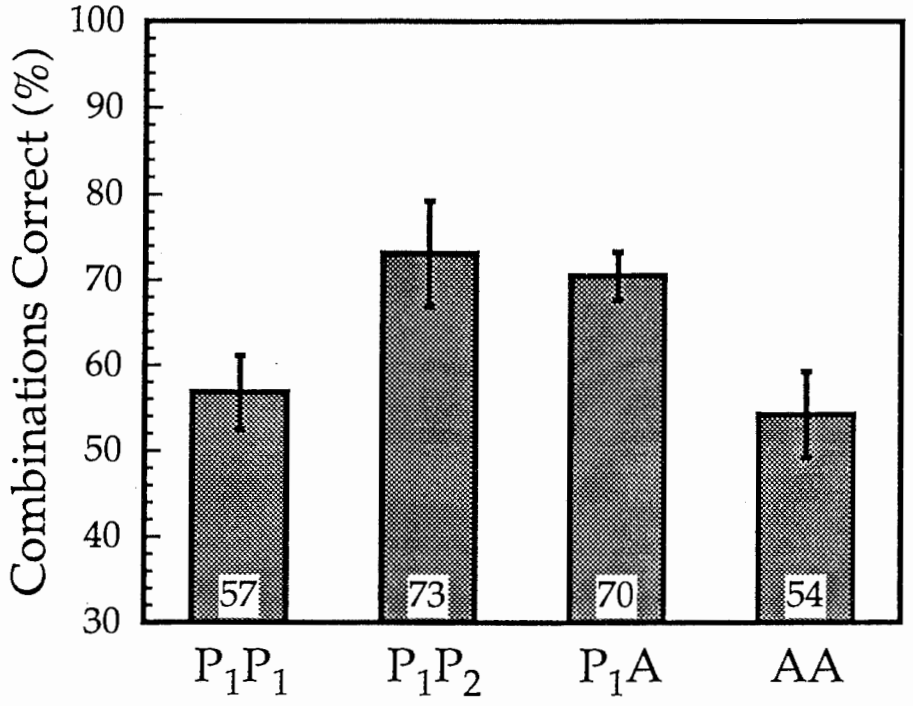

Figure 10. The results form Experiment 4 scored using the combinations-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

concurrent vowel was higher than for the P_1P_1 concurrent vowel ($p<0.01$). Accuracy of identification of the P_1A concurrent vowel was higher than the AA concurrent vowel ($p<0.01$), but not higher than the P_1P_1 concurrent vowel. Accuracy of identification of the P_1P_1 and AA concurrent vowels did not differ. Accuracy of identification of the P_1P_2 and P_1A concurrent vowels did not differ.

Figure 11 shows the results scored using the constituents-correct score averaged over the six listeners. A repeated measures ANOVA was performed with the factor of constituent type. Constituent type was found to be significant ($F_{5,25}=7.4$, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some constituent types differed from other constituent types. Accuracy of identification for the A/P_1 constituent type was higher than the A/A and P_1P_1 constituent types (both $p<0.01$). All other constituent types did not differ.

C. DISCUSSION OF EXPERIMENT 4

There were two procedural differences between Experiments 4 and 1 (number of tokens and the use of feedback). Despite these differences the pattern of results is very similar. For the results scored using the combinations-correct score both the P_1P_2 and the P_1A vowels are identified more accurately than both the P_1P_1 and the AA concurrent vowels, thus showing segregation has occurred. For the constituents-correct score the A/P_1 was identified more accurately than the A/A, but the P_1/A was not identified more accurately than the P_1/P_1. Thus, again suggesting that listeners can only segregate the whispered constituent and not the voiced constituent of the P_1A concurrent vowel.

This almost exact replication of the results of Experiment 1 suggest that neither the use of multiple tokens nor feedback produced the different patterns of results between Experiments 1 and 2.

The only difference between the two Experiments 1 and 2 which has not been investigated is that of matching the periodic and aperiodic vowel amplitudes. Experiment 5 seeks to find out how the periodic and aperiodic vowels should be matched by conducting a loudness matching experiment.

VI. EXPERIMENT 5

Experiment 5 uses the Japanese vowels to find how much more or less intense an aperiodic vowel had to be than a periodic vowel for it to be considered the same loudness by listeners.

A. STIMULI AND PROCEDURE OF EXPERIMENT 5

The same periodic and aperiodic single vowels that were used in Experiment 2 were used here, however, only the periodic vowels with a 100-Hz f0 were used. The vowels were nominally matched in terms of the dB RMS of their sample values.

The stimuli were presented on-line (Macintosh IIci with a Digidesign Sound Accelerator Card) and presented to both ear phones of a STAX SR-Λ headset. The presentation levels of the periodic vowels was 79 dB(A).

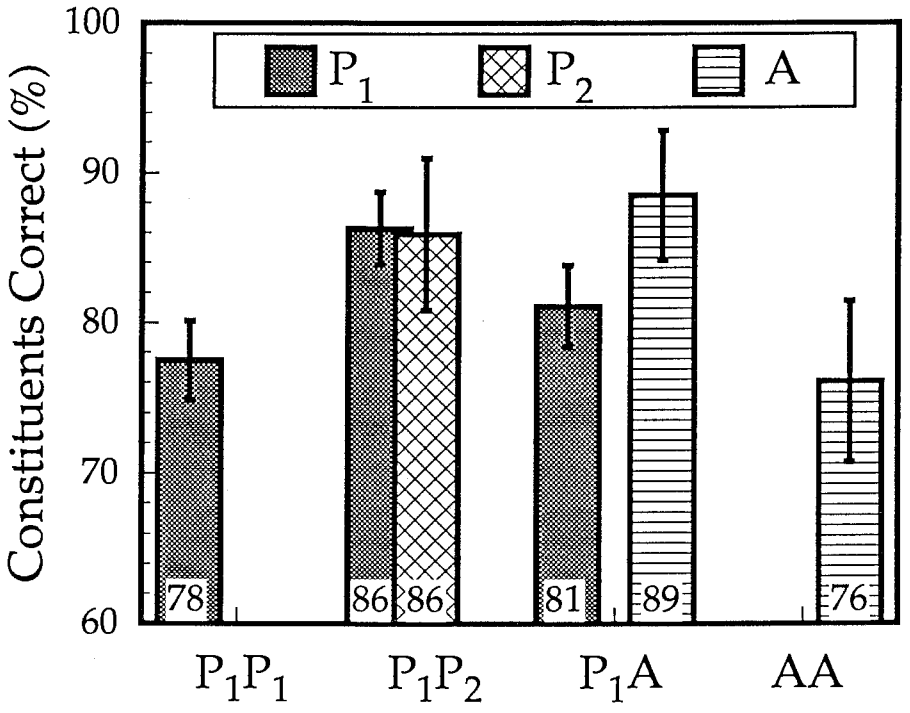

Figure 11. The results from Experiment 4 scored using the constituents-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the seven listeners and the different concurrent vowel combinations.

The experimental procedure was as follows: Each trial consisted of two intervals. In the first interval a periodic vowel was presented at a fixed intensity and in the second interval an aperiodic vowel was presented. The starting intensity of the aperiodic vowel was chosen randomly from within a +18 to -18 dB range of the amplitude of the periodic vowel in terms of dB RMS. The listeners' task was to adjust the intensity of the aperiodic vowel so that it was the same as the periodic vowel. They did this by pressing buttons on a computer screen such that the amplitude of the aperiodic vowel was either increased or decreased by 1, 2 or 3 dB. When the listeners considered the periodic and the aperiodic vowels were the same intensity they pressed another button. The intensity of the aperiodic vowel at this point was recorded and a new trial was begun.

Each of the five vowels were tested individually. Listeners completed twenty trials for each vowel of which the first four were counted as practice trials and were discarded from the results. The data for all five vowels was generally gathered in one experimental session.

All five listeners were Japanese female adults with normal hearing and were experienced with listening experiments. None of the five had taken part in Experiments 2 or 3.

No feedback was given to the listeners at any stage in this experiment. Also, the meaning of equal loudness was left intentionally vague when listeners were given experimental instructions.

B. RESULTS OF EXPERIMENT 5

The results were scored by computing the amplitude of the aperiodic vowel minus the amplitude of the periodic vowel for each of the sixteen trials in terms of dB RMS. This was done for each listener and for each vowel. The results are shown in Figure 12 averaged over the five listeners. The results were analyzed using a repeated measures ANOVA with the factor of vowel. Vowel was found to be not significant ($F_{4,16}$=0.2). The average of the listeners across the five vowels is 2.0 dB. A t-test was performed to see if the this figure differed from zero - the amplitude of the periodic vowels. It was significant (t_{24}=4.8, $p<0.01$).

C. DISCUSSION OF EXPERIMENT 5

The results of Experiment 5 show that listeners judged that the aperiodic vowels had to be 2.0 dB more intense than the periodic vowels for them to be considered equal in loudness. There was also no difference across the five vowels although there were individual listener differences.

The stimuli used in Experiment 1 were re-examined to discover whether the periodic vowels were louder than the aperiodic vowels or *vice versa*. It was found that on average the aperiodic vowels were 9.6 dB more intense than the 100-Hz f0 periodic vowels as measured by the dB RMS difference of the sample values. However, the aperiodic vowels from Experiment 1 contained intense low-frequency components which might affect this

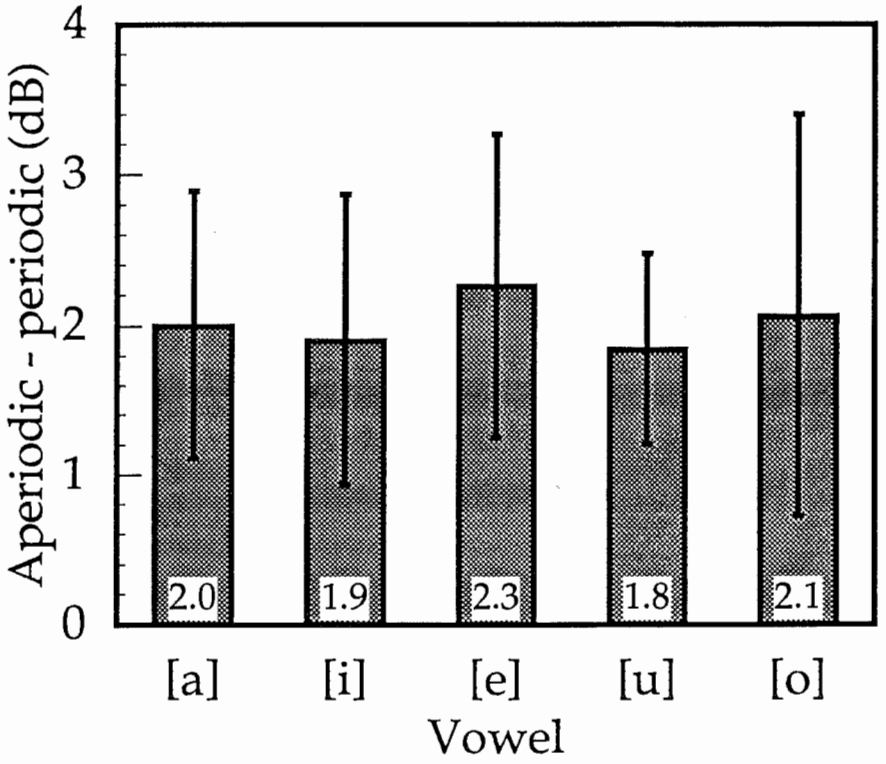

Figure 12. The results from Experiment 5 averaged over the five listeners. The figure plots the relative difference between the voiced and whispered vowels when the listeners consider the vowels to have equal loudness. The figure shows that the whispered vowels had to be 2 dB more intense in terms of dB RMS for them to be considered the same loudness as the voiced vowels.

figure. Therefore both the aperiodic vowels and the 100-Hz f0 periodic vowels were high-passed filtered with a cut-off frequency of 50 Hz. After this the difference was reduced to 8.8 dB.

In Experiment 2 the periodic and aperiodic vowels were equal in terms of dB RMS. From the results of Experiment 5 it can be seen that the method of equalizing dB RMS in Experiment 2 produced a much closer approximation to equal loudness than did matching using the excitation pattern in Experiment 1. Thus, the results of Experiment 2 are more likely to show listeners true ability to segregate periodic and aperiodic vowels than Experiment 1. However, in Experiment 2 the P_1/A constituent is 2 dB more intense than the A/P_1 constituent for them to be considered equally loud. Thus, it could be argued that the only reason the P_1/A constituent is segregated in Experiment 2 is that it is masking the A/P_1 constituent more than the A/P_1 constituent is masking the P_1/A constituent. Accordingly Experiment 6 was conducted to find out the constituents are segregated when the periodic and aperiodic vowels are at equal loudness.

VII. EXPERIMENT 6

Experiment 6 uses the Japanese vowels to determine how listeners segregated the periodic and aperiodic vowels when they were matched to be the same loudness using the results of Experiment 5.

A. STIMULI AND PROCEDURE OF EXPERIMENT 6

The same periodic and aperiodic single vowels that were used in Experiment 2 were also used here, however, the aperiodic vowels had been increased by 2 dB relative to the level of the periodic vowels according to the results of Experiment 5.

The stimuli were presented on-line (Macintosh IIci with a Digidesign Sound Accelerator Card) and presented to both ear phones of a STAX SR-Λ headset. The presentation levels of the periodic vowels was the same as for Experiment 2 at 78.8 dB(A).

A set of 10 listeners were used of which one was the first author and the other nine where Japanese female listeners who were familiar with listening experiments. The listeners were first trained with a non-random sequence of 5 repetitions of the 3 types (i.e. P_1, P_2 and A types) of single vowels making 75 stimuli in total (5 repetitions x 3 types x 5 vowels). Second, a randomized sequence of the single vowels was presented to listeners to test their ability to identify the single vowels, this consisted of 10 repetitions of the 3 types of single vowels, making 150 stimuli in total (10 repetitions x 3 types x 5 vowels). Thirdly, a randomized sequence of the four types of concurrent vowels (i.e. P_1P_1, P_1P_2, P_1A and AA types) was presented to listeners, this consisted of 10 repetitions of the 20 concurrent vowel combinations making 800 stimuli in total (10 repetitions x 20 concurrent vowel combinations x 4 concurrent vowel types).

C. RESULTS OF EXPERIMENT 6

For the single vowels, the average identification of the 10 listeners was 95.6%, 96.0% and 86.6% correct for the P_1, P_2 and A vowels respectively. A repeated measures ANOVA

was performed with the factor of vowel type. Vowel type was found to be significant ($F_{2,18}$=4.17, $p<0.05$). No pair-wise *post-hoc* Scheffé test were significant, but contrast tests showed that the A vowel was identified less well than both the P_1 and P_2 vowels ($F_{2,18}$=7.2, $p<0.05$).

Again the results were scored using both the combinations-correct and constituents-correct scores (see the results section of Experiment 1). The results of the combinations-correct score are presented in Figure 13 averaged over the six listeners. A repeated measures ANOVA was performed with the factor of concurrent vowel type. Concurrent vowel type was found to be significant ($F_{3,27}$=28.8, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some concurrent vowel combinations differed from other concurrent vowel combinations. Accuracy of identification of the P_1P_2 concurrent vowel was higher than for the P_1P_1 concurrent vowel ($p<0.01$). Accuracy of identification of the P_1A concurrent vowel was higher than the AA concurrent vowel ($p<0.01$) and higher than the P_1P_1 concurrent vowel ($p<0.05$). Accuracy of identification of the P_1P_1 and AA concurrent vowels did not differ. Accuracy of identification of the P_1P_2 and P_1A concurrent vowels did not differ.

Figure 14 shows the results scored using the constituents-correct score averaged over the six listeners. A repeated measures ANOVA was performed with the factor of constituent type. Constituent type was found to be significant ($F_{5,45}$=20.8, $p<0.01$). Pair-wise *post-hoc* Scheffé tests were performed to test whether the accuracy of some constituent types differed from other constituent types. Accuracy of identification for both the P_1/P_2 and P_2/P_1 constituent types were higher than the P_1P_1 constituent type (both $p<0.01$). Accuracy of identification for the P_1/A constituent type was higher than for the P_1/P_1 constituent type ($p<0.01$). Accuracy of identification for the A/P_1 constituent type was higher than for the A/A constituent type ($p<0.05$). Accuracy of identification for the P_1/P_1 and A/A constituent types did not differ.

D. DISCUSSION OF EXPERIMENT 6

The results of Experiment 6 show that listeners can segregate both the P_1/A and the A/P_1 constituent types when they have been matched for equal loudness. Thus, this confirms the hypothesis that listeners can segregate both the periodic and aperiodic constituents of the P_1A concurrent vowels.

VIII. GENERAL DISCUSSION

In the first section of the General Discussion we will summarize the results of the experiments. In the second section we will discussion how listeners might be segregating the concurrent vowel stimuli.

A. SUMMARY OF EXPERIMENTAL RESULTS

Experiments 1, 2, 3, 4 and 6 showed that listeners could use the difference in voicing between aperiodic and periodic vowels to segregate them. These experiments also showed

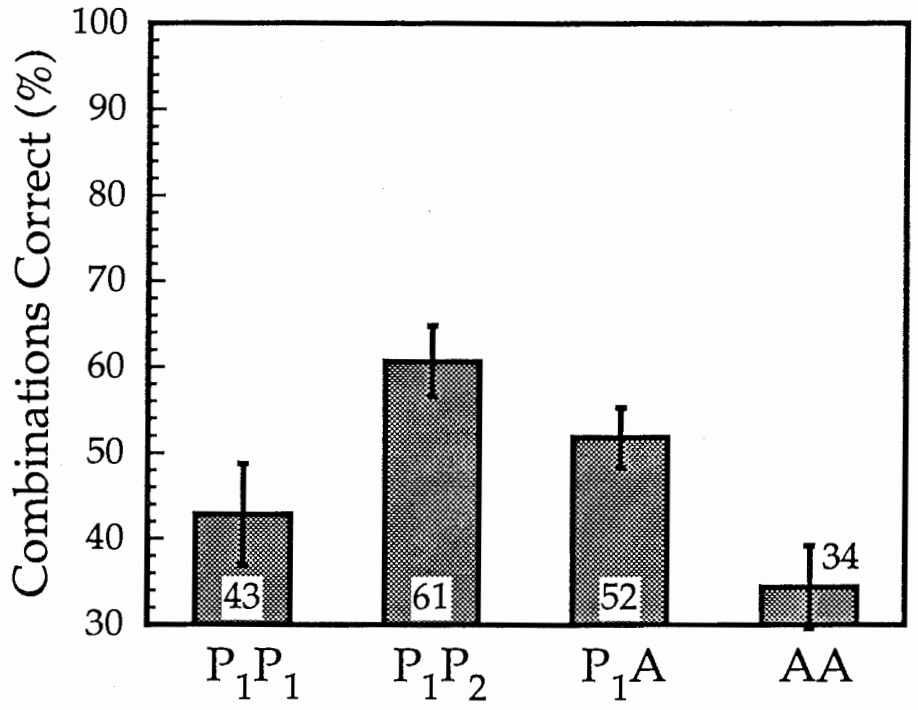

Figure 13. The results form Experiment 6 scored using the combinations-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the ten listeners and the different concurrent vowel combinations.

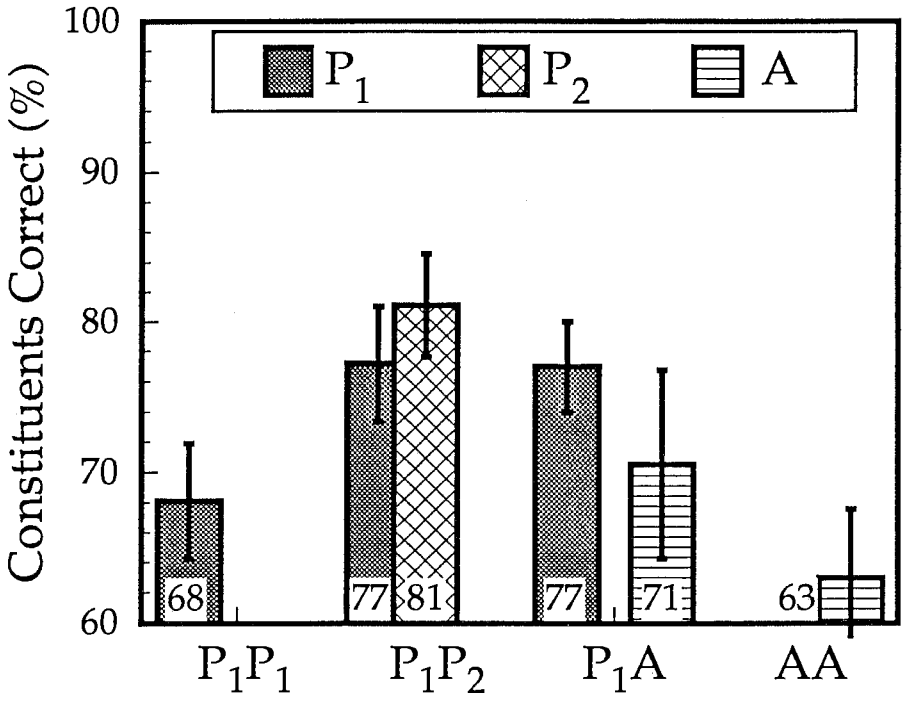

Figure 14. The results from Experiment 6 scored using the constituents-correct score. Error bars plot +/- one intra-listener standard deviation about the mean. The results were averaged over the ten listeners and the different concurrent vowel combinations.

that the ability of listeners to use this difference in voicing was as good as a difference in pitch between two voiced vowels.

Experiment 1 using vowels matched for equal loudness by their excitation patterns suggested that only the aperiodic constituent of the periodic/aperiodic concurrent vowel was segregated by listeners. However, Experiment 2 using vowels matched by equal RMS suggested that both constituents of the periodic/aperiodic concurrent vowels could be segregated by listeners. Experiments 3 and 4 showed that the procedural differences between Experiments 1 and 2 of the use of feedback, multiple vowel tokens and amplitude differences between vowels did not explain the difference in results between the experiments. One difference between Experiment 1 and 2 was the matching procedure used. In Experiment 1 the vowels were matched for "equal" intensity using the excitation pattern representation. However, due to an error in matching procedure the aperiodic vowels were 9.6 dB more intense than the periodic vowels in Experiment 1. The vowels in Experiment 2 were matched by the RMS dB of the sample values.

To find the most appropriate matching procedure Experiment 5 was conducted to match the periodic and aperiodic vowels to be equally loud for listeners. The results of the experiment show that the aperiodic vowels had to be 2 dB more intense than the periodic vowels in terms of RMS dB of the sample values. Finally Experiment 6 was conducted to find how the periodic/aperiodic concurrent vowels were segregated when the vowels had been matched for equal loudness using the results of Experiment 5. The results of Experiment 6 show that both the periodic and aperiodic constituents of the periodic/aperiodic concurrent vowels were segregated from each other.

In Experiments 1, 4 and 6 listeners were able to identify the concurrent vowels in which both constituents were both aperiodic as well as the concurrent vowels in which both constituents were periodic with no f0 difference. However, this was not the case in Experiments 2 and 3. Scheffers (1983) found that his listeners were much worse at identifying the concurrent vowels in which both constituents were aperiodic than when both were periodic with no f0 difference. There are a number of procedural difference between Scheffers experiment and those described here and indeed within the experiments described here.

In Experiments 1 and 6 listeners were presented with the three types of single vowels in isolation and the identification rates were measured. For Experiment 1 listeners identified 99% of both the periodic and aperiodic single vowels. However, for Experiment 6 listeners identified only 87% of the aperiodic vowels and 96% of the periodic single vowels. Even though listeners did not significantly perform worse for the concurrent vowels in which both constituents were periodic with no pitch difference, than for the concurrent vowels in which both constituents were aperiodic in Experiment 6, these identification rates for the single vowels suggest that for some reason the identifiability of the English aperiodic vowels and the Japanese aperiodic vowels varies. The difference in

matching procedure between the vowel sets ought not have affected the results as single vowel identification does not vary much with amplitude.

There are number of differences between the English and Japanese aperiodic vowels. These are: formant frequencies, number of formants and synthesis methods. Assuming that the two different versions of the Klatt (1980) synthesizer worked as they were created to[1], either the difference in formant frequencies or the different number of formants resulted in the change in single vowel identifiability between Experiment 1 and 6. The number of formants, three versus five, should not effect the identifiability of the aperiodic vowels as generally only the first two and occasionally first three formants contribute to vowel quality (Joos, 1948; Delattre *et al*, 1952). Thus, the most probable explanation is that because of the formant frequencies chosen, the aperiodic Japanese vowels were more difficult for listeners to identify compared to the periodic Japanese vowels and both periodic and aperiodic English vowels.

This difference in identifiability between the Japanese periodic and aperiodic vowels in Experiments 2, 3 and 6 should not have affected how listeners segregated the periodic/aperiodic concurrent vowels. This is because it was inferred whether segregation occurred or not by comparing the identification levels of the constituents of the periodic/aperiodic concurrent vowel with the identification levels of the reference concurrent vowels in which both constituents were either periodic or aperiodic. Thus, this canceled the different identifiabilities of the Japanese aperiodic and periodic vowels.

B. SEGREGATING CONCURRENT VOWELS

There are two different ways of looking at how listeners might segregate the concurrent vowels in which one constituent is periodic and the other is aperiodic and consequently other concurrent vowels and other sounds which overlap in time.

One way is to approach the problem by assuming that one vowel is the target constituent and the other is the interfering constituent and then to examine how the target constituent is segregated. Each of the concurrent vowel constituents is therefore both the target constituent and the interfering constituent in turn. A number of studies have used this approach to examine the segregation of double vowels (Lea, 1992; Lea and Tsuzaki, 1993; de Cheveigné, 1993).

Another approach to the problem is to assume that both constituents of the concurrent vowels are target constituents and that both vowel streams are formed simultaneously. This approach is based on the work of Bregman (1990).

These two approaches lead to different interpretations about how listeners are segregating the concurrent vowels. Therefore, we will outline how listeners might segregate concurrent vowels with each approach and then discuss the merits of the two approaches.

THE TARGET AND INTERFERER APPROACH

According to the literature on voice-separation algorithms (see de Cheveigné, 1993 for a recent review) there are two strategies for segregating competing voices.

The spectro-temporal structure of the target vowel can be used to select or enhance the components of the target vowel and hence segregate it from the interfering vowel. This has been termed the "enhancement" strategy by de Cheveigné (1993), but we will use the more general term of the "selection" strategy (Lea, 1992; Lea and Tsuzaki, 1993). For the selection strategy to be useful the target vowel must have some property which distinguishes it from the interfering vowel which is useful for segregation (Stubbs and Summerfield, 1990; Lea, 1992).

The spectro-temporal structure of the interfering vowel can be used to remove the components of the interfering vowel and hence to segregate the target vowel. This has been termed the "cancellation" strategy by de Cheveigné (1993) which is the term we will use here. For the cancellation strategy to be useful the interfering vowel must have some property which distinguishes it from the target vowel which is useful for segregation (Stubbs and Summerfield, 1990; Lea, 1992).

Both strategies have disadvantages. The selection strategy should work best when the signal to noise ratio (SNR) of the target voice is greater than 0 dB (Stubbs and Summerfield, 1990), whereas the cancellation strategy should work best when the SNR of the target voice is less than 0 dB (Hanson and Wong, 1983, 1984). If the target voice is aperiodic, i.e. it has no structure, then it can only be segregated from an interfering voice by using the structure of the interfering voice. Also, if the interfering voice is aperiodic, i.e. it has no structure, then the target voice can only be segregated by using the structure of the target voice.

Accordingly to get the greatest possible amount of segregation both strategies of segregation should be used simultaneously where possible. Two studies of speech segregation have tried this hybrid approach (Stubbs and Summerfield, 1990, 1991; Meddis and Hewitt, 1992).

Since the periodic constituent of the periodic and aperiodic concurrent vowels is segregated we can infer that listeners are segregating it by using the selection strategy. Listeners cannot use the cancellation strategy as the aperiodic constituent has no structure which can be used to cancel it. Since the aperiodic constituent of the periodic and aperiodic concurrent vowels is segregated we can infer that listeners are using the subtraction strategy to cancel the periodic vowel and hence segregate the aperiodic vowel. Listeners cannot be using the selection strategy as the aperiodic vowel has no structure for selection to work.

Therefore according to the results of the perceptual experiments listeners appear to use both selection and cancellation where appropriate to segregate the periodic and aperiodic concurrent vowels. Thus, listeners are using a hybrid approach.

THE TWO TARGET APPROACH

In the natural environment sounds from two or more different sources sometimes overlap in time and sometimes do not overlap in time. The segregation or integration of sounds that do not overlap in time and must be separated into two or more sources is called sequential grouping by Bregman (1990). The segregation or integration of sounds that do overlap in time and must be separated into two or more sources is called simultaneous grouping by Bregman (1990).

Usually in the sequential grouping of sounds, energy in the internal auditory spectrogram can only provide evidence of a single auditory stream (Bregman, 1990). This is called the principle of exclusive allocation, as energy in the internal auditory spectrogram can belong to one stream or another stream, but not both.

However, in simultaneous grouping it is often the case that energy in the internal auditory spectrogram has been produced by more than one physical source. A concurrent vowel made up of two periodic vowels with the same f0 is an example of this. Harmonics from both vowels have the same frequency, start and stop simultaneously and have the same phase and yet it is possible for listeners to identify both vowels with an accuracy far greater than chance level. Here the principle of exclusive allocation is being broken.

The exclusive allocation rule sometimes breaks down for sequential grouping of sounds. A famous example of this is the "duplex perception of speech" (see Bregman (1990) and Liberman (1982) for reviews). Duplex perception does no just occur with speech, but also occurs in music (Pastore *et al*, 1983; Collins, 1985), with creaking doors (Fowler and Rosenblum, 1988, reported by Bregman, 1990) and pure tone stimuli (Steiger, 1983, reported by Bregman, 1990).

To account for examples which break the exclusive allocation rule Bregman (1990) formulated a new rule – the constraint of consistency or noncontradiction. This constraint says that the exclusive allocation rule should be obeyed unless there is evidence that more than one interpretation is possible and when there is evidence of more than one interpretation, the energy should be divided according to the evidence.

Another problem is what happens when one stream is segregated from a mixture. For example if we have two events P and Q, what happens when P is segregated from the mixture PQ? Does forming the stream P mean that a residue stream of Q results? Bregman (1990) answers this question by saying that yes this often the case, but that it is not necessarily so, although he does not give any examples of cases where this is not the case. When P is removed from PQ and a residue of Q results then Q is segregated even though no property of Q might exist to remove Q from P. However, there must be some reason for grouping together the residue after removing P from PQ (Bregman, 1990).

So, how are the periodic and aperiodic concurrent vowels segregated if both vowels are target vowels? We still have the fact that only the periodic constituent has structure, therefore the components of the periodic constituent must be grouped together to form a stream. Due to the constraint of noncontradiction there is no reason to suppose that the

energy of the periodic vowel is shared with any other source in the internal spectrogram, therefore this energy is largely removed. Once the periodic vowel has been removed then the residue must be grouped together to form a stream for the aperiodic constituent. There is evidence that the remaining energy in the internal auditory spectrogram should be grouped together as all the remaining energy has the same onset and offset times. Thus, streams for both the periodic and aperiodic constituents have been formed and both have been segregated.

COMPARING THE TWO APPROACHES

The two different approaches to the problem address the problem of concurrent vowel segregation from different directions. However, we would like to show that although their conclusions are the same although the two target approach is both a richer and more accurate description of the perceptual processes involved.

The description of the target and interferer approach above concludes by saying both selection and cancellation must be occurring for the periodic and aperiodic constituents of the periodic/aperiodic concurrent vowels to be segregated. When the aperiodic constituent is the target it is segregated by the cancellation of the periodic constituent. However, before this can happen the periodic constituent must first be selected before it can be canceled. Thus, the segregation of the aperiodic constituent target implicitly requires that the periodic constituent must also be a target. Hence the target and interferer approach can be viewed as the same as the two target approach except for the terminology.

Not only does the two target approach provide a more accurate description of what is occurring it uses the language of psychology where as the target and interferer uses the language of speech separation algorithms. Thus, on both accounts the two target approach to concurrent segregation is more desirable.

This two target approach to concurrent vowel segregation does not imply that in the natural environment that the auditory system is capable of only forming two streams. The aim of the auditory system is to provide a complete a description of the auditory scene as possible. Thus, the auditory system must try and form as many streams as it can identify sound sources. Listeners can then focus on one or two of the auditory streams to hear them consciously. If a stream cannot be formed for a sound source the sound will not be identifiable to the listener and will have been masked.

The strategies of selection and cancellation are implicit in the general formation of streams. To form the periodic constituent stream, selection is used. Due to the constraint of noncontradiction the energy of the periodic constituent is then canceled from the internal auditory spectrogram to form the stream of the aperiodic constituent. In fact the strategy of cancellation is very important in the formation of auditory streams. Any energy in the internal spectrogram that is not shared between streams must be canceled after each stream has been formed.

Auditory scene analysis can be thought of as a competitive process in which streams use rules to fight for energy in the internal spectrogram, the stream which has the strongest evidence for including the energy "wins" the energy and unless there is evidence for including the energy in another stream, then the energy is denied from other streams by canceling it from the internal auditory spectrogram. Such a strategy has been used in early models of auditory scene analysis (Cooke, 1991; Brown, 1992; Ellis, 1993).

IX. ACKNOWLEDGMENTS

Experiment 1 was performed at the Institute of Hearing Research while author APL was supported by a studentship from the Medical Research Council. The other experiments were performed at ATR. Thanks to Hideki Kawahara who commented on an earlier version of this paper.

---

[1]The formant frequencies, formant bandwidths and noise distribution were examined for the aperiodic vowels synthesized by the two methods. Nothing extraordinary was found, vowels produced by both synthesis methods had the correct formant frequencies, formant bandwidths and the sample values had Normal distributions.

X. APPENDIX A

The aim of the iterative matching procedure was to equate the excitation patterns of periodic and aperiodic vowel tokens at the harmonic frequencies of the periodic token.

A first estimate of harmonic amplitudes for the voiced token was obtained from the excitation pattern of the aperiodic vowel at the harmonic frequencies of the periodic token by Equation 1. In the equation, P_i specifies the amplitude in dB of the ith harmonic of the periodic token at synthesis and A_i is the amplitude in dB of the ith harmonic of the excitation pattern of the aperiodic token:

(1) $P_i = A_i - 60$

If P_i was greater than 45, it was set to 45 and if P_i was less than or equal to zero, it was set to be just greater than 0. All amplitudes were computed in dB.

These estimated amplitudes were used to control a harmonic synthesizer. Once the periodic token was synthesized its excitation pattern was computed and compared to the excitation pattern of the aperiodic token. The error between the two excitation patterns was used to revise the estimated harmonic amplitudes values according to Equation 2. L_i is the amplitude of the ith harmonic of the latest estimate of the periodic vowel and E_i is the amplitude of the error at the ith harmonic:

(2) $P_i \Leftarrow P_i + E_i \qquad E_i = A_i - Li$

If the absolute value of the error, E_i was greater than 10 dB, then the error was set to 10 dB with the appropriate sign. This was done to avoid fluctuations between large positive and negative errors. If P_i was less than or equal to zero, it was set to be just greater than zero.

This iterative process of synthesis and comparison continued for 20 iterations, or until the average error for all harmonics was less than 0.033 dB and the largest error for any harmonic was less than 0.33 dB. After 20 iterations if both criterion were not met, the procedure was discontinued as little improvement was found if more iterations were performed. On the whole the procedure produced an excellent match between the voiced and whispered vowel excitation patterns, with errors rarely exceeding 1 dB at any harmonic.

XI. REFERENCES

ANSI (1969). "Specifications for Audiometers", American National Standards Institute.

Assmann, P. F., & Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies", J. Acoust. Soc. Am., 88, 680-697.

Bregman, A. S. (1990). "Auditory Scene Analysis", MIT Press: Cambridge.

Broadbent, D. E., & Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. 29, 708-710.

Brokx, J. P. L., & Nooteboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices", J. of Phonetics, 10, 23-36.

Brokx, J. P. L., Nooteboom, S. G., & Cohen, A. (1979). "Pitch difference and the intelligibility of speech masked by speech", (14 55-60). IPO.

Brown, G. J. (1992). "Computational auditory scene analysis: a representational approach", Doctoral Thesis, University of Sheffield.

Chalikia, M., & Bregman, A. (1989). "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation," Percept. Psychophys. 46, 487-496.

Cherry, E. C. (1953). "Some experiments on the recognition of speech with one and two ears", J. Acoust. Soc. Am., 23, 975-979.

Collins, S. (1985). "Duplex perception of musical stimuli: A further investigation", Perception and Psychophysics, 38, 172-177.

Cooke, M. P. (1991). "Modeling auditory scene processing and organisation", Doctoral Thesis, University of Sheffield.

Culling, J. F., & Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by f0", J. Acoust. Soc. Am., 93, 3454-3467.

Darwin, C. J. (1981), "Perceptual grouping of speech components differing in fundamental frequency and onset-time", Q. J. Exp. Psychol. 33A, 185-207.

de Cheveigné, A (1993) "Separation of concurrent harmonic sounds: Fundamental frequency estimation, and a time-domain cancellation model of auditory processing", J. Acoust. Soc. Am., 93, 3270-3290.

Delattre, P., Liberman, A. M., Cooper, F. S., and Gertsman, L. J. (1952) "An experimental study of the determinants of vowel color: observations on one- and two- formant vowels synthesized from spectrographic patterns", Word, 8, 195-210.

Ellis, D. P. W. (1993). "A computer implementation of psychoacoustic grouping rules", J. Acoust. Soc. Am., 93, 2308.

Fowler, C. A. and Rosenblum, L. D. (1988). "The perception of phonetic gestures", Presentation in the conference, Modularity and the Motor Theory of Speech Perception, Haskins Laboratories, New Haven, Conn., June 5-8, 1988.

Hanson, B. A., & Wong, D. Y. (1983). "Processing techniques for intelligibility improvement to speech with co-channel interference", (83 225). Rome Air Development Center.

Hanson, B. A., & Wong, D. Y. (1984). "The harmonic magnitude suppression (HMS) technique for intelligibility enhancement in the presence of interfering speech", Proceedings IEEE ICASSP (pp. 18A.5.1-18A.5.4).

Helmholtz, H. L. F. von (1863). Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. First edition. F. Vieweg, Braunschweig.

Joos, M. (1948). "Acoustic phonetics", Lang., 24(Suppl.), 1-36.

Klatt, D. H. (1980). "Software for cascade/parallel formant synthesizer", J. Acoust. Soc. Am., 67, 971-995.

Lea, A. P. (1992) "Auditory Modeling of vowel perception", unpublished Doctoral Thesis, University of Nottingham.

Lea, A. P. and Tsuzaki, M. (1992). " Segregation of whispered and voiced vowels in English and Japanese", J. Acoust. Soc. Am., 93, 2403.

Liberman, A. M. (1982). "On finding that speech is special", American Psychologist, 37, 148-167.

McAdams, S. E. (1989), "Segregation of concurrent sounds. I: effects of frequency modulation coherence", J. Acoust. Soc. Am., 86, 2148-2159.

Meddis, R., & Hewitt, M. (1992). "Modeling the identification of concurrent vowels with different fundamental frequencies", J. Acoust. Soc. Am., 91, 233-245.

Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter shapes and excitation patterns," J. Acoust. Soc. Am. 74, 750-753.

Moore, B. C. J., and Glasberg, B. R. (1986). "The role of frequency selectivity in the perception of loudness, pitch and time", In B. C. J. Moore (Ed.), Frequency Selectivity in Hearing, Academic: London. pp. 250-308.

Moore, B. C. J., and Glasberg, B. R. (1987). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hearing Research 28, 209-225.

Moore, B. C. J., and O'Loughlin, B. J. (1986). "The use of nonsimultaneous masking to measure frequency selectivity and suppression", In B. C. J. Moore (Ed.), Frequency Selectivity in Hearing. Academic: London. pp. 179-249.

Palmer, A. R. (1990). "The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea-pig cochlear-nerve fibers", J. Acoust. Soc. Am., 88, 1412-1426.

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., and Szczesiul, R. (1983). "Duplex perception with musical stimuli", Perception and Psychophysics, 33, 469-474.

Patterson, R. D. (1974). "Auditory filter shape", J. Acoust. Soc. Am. 55, 802-809.

Patterson, R. D., and Holdsworth, J. (1990). "An introduction to auditory sensation processing", Applied Psychology Unit, Cambridge.

Patterson, R. D., and 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, Academic: London. pp. 123-177.

Patterson, R. D., Holdsworth, J., and Allerhand M. (1992). "Auditory models as preprocessors for speech recognition", In M. E. H. Schouten (ed.), The Auditory Processing of Speech: From the Auditory Periphery to words; Mouton de Gruyter: Berlin, 67-83.

Scheffers, M. T. M. (1983). "Sifting vowels: auditory pitch analysis and sound segregation", Doctoral Thesis, University of Groningen.

Steiger, H. (1983). "Influences of sequential organization processes on the use of binaural cues", Unpublished Doctoral Thesis, McGill University.

Stubbs, R. J., & Summerfield, Q. (1988). "Evaluation of two voice-separation algorithms using normal-hearing and hearing-impaired listeners", J. Acoust. Soc. Am., 84, 1236-1249.

Stubbs, R. J., & Summerfield, Q. (1990). "Algorithms for separating the speech of interfering talkers: evaluations with voiced sentences, and normal-hearing and hearing-impaired listeners", J. Acoust. Soc. Am., 87, 359-372.

Stubbs, R. J., & Summerfield, Q. (1991). "Effects of signal-to-noise ratio, signal periodicity, and degree of hearing impairment on the performance of voice separation algorithms", J. Acoust. Soc. Am., 89, 1383-1393.

Summerfield, Q., & Assmann, P. F. (1991). "Perception of concurrent vowels: effects of pitch-pulse asynchrony and harmonic misalignment", J. Acoust. Soc. Am., 89, 1364-1377.

von Bekesy, G. (1960). "Experiments in Hearing", Translated by Weaver, E.G. New York: McGraw-Hill.

Winer, B. J., Brown, D. R. and Michels, K. M. (1991). "Statistical principles in experimental design," (3rd edition), McGraw-Hill, New York.

Zwicker, E. and Feldtkeller, R. (1967). Das Ohr als Nachrichtenempfänger. Hirzel Verlag: Stuttgart.

Zwicker, U. T. (1984). "Auditory recognition of diotic and dichotic vowel pairs", Speech Comm., 3, 265-277.