A PHONETIC ANALYSIS OF THE INDONESIAN VOWEL SYSTEM: A PRELIMINARY ACOUSTIC STUDY [1]

Ellen van Zanten & Vincent J. Van Heuven

Leyden University

Formant structure and duration of the six Indonesian monophthongs spoken in isolation and in monosyllables were analysed. Of the vowels spoken in isolation with a rise-fall-pattern fundamental frequency was also measured. The results reveal amongst other things that there is very little overlap in the F_1/F_2-plane between the realisations of the six vowels in context. Secondly, the "pepet" /ə/, which is traditionally regarded as a central vowel, appears to be a more closed vowel when spoken in context. As regards its fundamental frequency /ə/ also seems to agree with the closed vowels /i/ and /u/. The measurements on duration indicate for the vowels in context a rough partition into closed vowels on the one hand and mid-vowels and open vowels on the other; the durations of the vowels spoken in isolation do not seem to differ significantly.

1. INTRODUCTION

1.1 AIM OF THIS STUDY

The first aim of the experiment described in this paper is to contribute to the experimental phonetic description of the Indonesian sound system. Secondly, the results of the measurements will be used for our forthcoming perceptual study of the possibly different ways in which Indonesian speakers with different regional backgrounds divide the vowel space (van Zanten and van Heuven, 1983).

Acoustic measurements on Indonesian vowels have, as far as we are aware, only been reported by Zubkova (1976). She determined the center frequencies of the first four formants of the Indonesian monophthongs, each spoken once by one female and four male speakers in a number of conditions, viz. as isolated vowels, embedded in different positions and different consonantal environments in monosyllabic and polysyllabic words, as well as under different accent conditions. In the presentation of her results, however, Zubkova does not differentiate for the various contexts. In the present study we have tried to remedy this problem by strictly separating vowels spoken in isolation from vowels spoken in context. We also tried to keep the segmental environment of the vowels in the latter category as constant as possible. Moreover, several realisations per speaker were collected in order to get an impression of both inter- and intra-speaker variability.

1.2. OPERATIONALISATION

The study is restricted to the Indonesian monophthongs, leaving aside the category of diphthongs. Indonesian grammars indicate six monophthongs for the Indonesian vowel system, viz. /i, e, a, o, u, ə/. There is allophonic variation between realisations in open and in closed syllables, but there is no agreement on the extent to which allophonic variation takes place. Amran Halim (1974) for instance, who describes Indonesian as it is spoken in South Sumatra, states that each of the six vowels, except /a/ and /ə/, is phonetically represented by two allophonic variants; according to Ross Macdonald and Soenjono Dardjowidjojo (1967) allophonic variation occurs in all monophthongs except /ə/, according to Teeuw (1978) only in /i, e, a, o/, and according to Arsath Ro'is (1980) only in /i/ and /a/, whereas /e/ and /o/ - apart from a few exceptions - "are pronounced as /ɛ/ and /ɔ/ in Dutch".

In our experiment the six monophthongs were spoken by male native speakers of Indonesian. To get as close as possible to real speech the vowels were embedded in words which in turn were incorporated in sentences. Due to phonotactic constraints it is not always possible to keep the context identical for all vowels. In view of naturalness, however, we did not take recourse to nonsense words. The speakers were also asked to pronounce the vowels in isolation, in order to be able to obtain a better estimate of their target values.

2. METHOD

2.1. RECORDINGS

The six Indonesian monophthongs were

recorded in isolation as well as embedded in mono- and disyllabic words. The stimulus words were selected such that the context for all six vowels was, as far as possible, the same; all stimuli had the form CVC(VC), where C was always a voiceless plosive. Voiceless plosives were chosen because they are relatively easy to segment in a visual representation of the speech signal. Because of phonotactic constraints and lexical gaps it was not possible to find a context CVC(VC) for /ə/ and /e/; here the words *ke* 'towards' /kə/, *Tebet* 'quarter of Jakarta' /təbət/, *pes* 'plague' /pes/ and *tetes* 'drop' /tetes/ were chosen.

All stimulus words were spoken - in pre-pausal position - embedded in the sentence: "*Ini kata* ...", 'This is the word ...', with the purpose to keep the prosody as identical as possible. Similarly, the vowels spoken in isolation were embedded at the end of a fixed sentence frame: "*Dalam kata* ... *terdapat bunyi* ...", 'In the word ... we find the sound ...'.

a. *Procedure*

The stimulus words in their carrier-sentences were typed on individual cards and presented five times to the subjects in the same random order. Immediately after this the vowels in isolation were pronounced once by each subject. The speakers were asked to read each sentence on their own, and to observe a pause between consecutive sentences. (For a list of the complete stimulus material see Appendix 1.)

The recordings were made in a sound-proofed studio of the Indonesian Broadcasting Corporation onto a Nagra IV S taperecorder (19 cm/s) with an AKG-CK 8 hypercardioid microphone.

b. *Informants*

Ten speakers took part in the experiment. All of them were male adults, middle class and at the time living in Jakarta. Four were of Javanese descent and three of Chinese. Two of the speakers were "real" Jakartans ("Betawi asli") and one called himself "campuran": he was born in Ujung Pandang and had lived in Central Java before he came to Jakarta.

2.2. MEASUREMENTS

In the presentation of the results we shall limit ourselves to the analysis of vowels spoken in isolation and in monosyllables only. The following acoustic characteristics of the vowels were measured:

a. *Formant Structure*

As is customary in an acoustic description of vowel systems the center frequencies of the lower formants - i.e. groups of adjacent overtones amplified by the resonance characteristics of the vocal tract - were measured. The frequency of the lowest formant, F_1, is inversely related to vowel height; the second formant, f_2, (or rather the difference between the frequencies of F_2 and F_1) corresponds to the degree of backness in a traditional vowel diagram (cf. Ladefoged, 1975, 173). F_1, F_2, and to a lesser extent F_3 are decisive in distinguishing vowels from each other. The center frequencies of these three formants were estimated from narrowband section spectrograms (Kay Sonagraph 6061, filter bandwidth 50 Hz) taken at a point in time halfway through the vowel. In the scale used 1 mm along the frequency-axis represents 82 Hz.

b. *Duration*

Durations of the vowels in monosyllables were measured from broad-band spectrograms (Kay Sonagraph 6061, filter bandwidth 300 Hz; 1 mm along the time-axis corresponds to 7.6 ms). As the vowel /ə/ in the word *ke* was not closed in a clear way by a final consonant it was decided not to measure duration in context of this vowel.

For the vowels in isolation a more detailed temporal analysis was made from oscillograms (Honeywell Visicorder 2206, 20 cm/sec) and a parallel intensity trace (Frøkjaer Jensen IM 360, 20 ms integration time, full bandwidth). The duration of the vowel was defined as the time interval between the moments in which the intensity of the vowel exceeded resp. dropped below 10% (along a decibel-scale) of its peak value (Debrock, 1977). As rise time the time segment was measured during which the intensity increased from 10% to 90% of its peak value (t_0 and t_1, resp.), and as decay time the interval between the moments where the intensity dropped from 90% to 10% (t_2 and t_3, resp.). The steady time then equals the overall vowel duration minus its rise and decay time. The moments t_0 and t_2 were chosen as late as possible, t_1 and t_3 as early as possible (see also Figure 1).

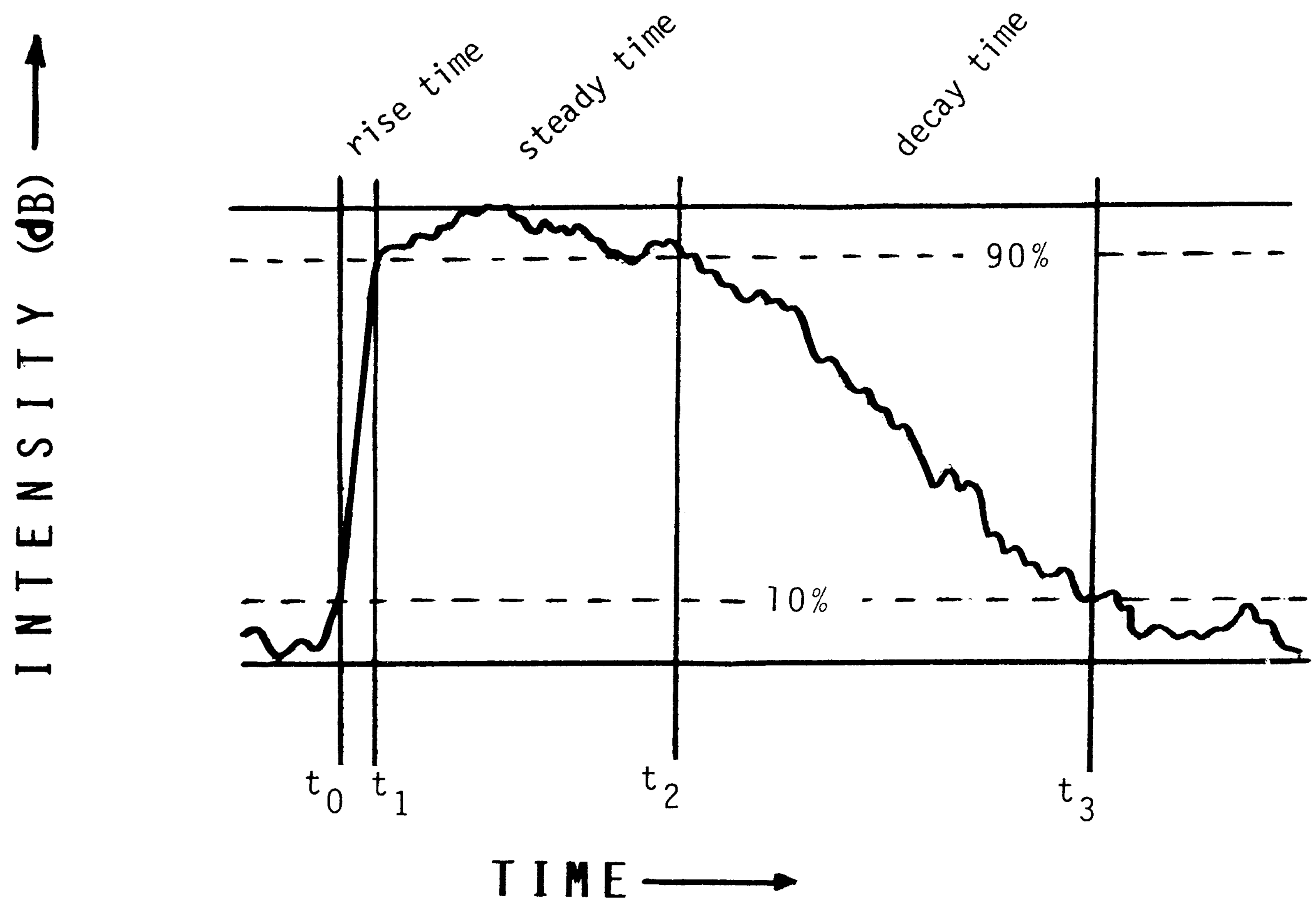

Figure 1: an example of an intensity trace of a vowel spoken in isolation.

c. *Periodicity (F_0)*

Fundamental frequency was only measured for the vowels spoken in isolation, on the basis of an F_0-trace written on a third channel on the oscillogram (Frøkjaer-Jensen FFM 650 F_0-meter). A logarithmic scale was chosen in which 10 mm equalled a fixed musical interval of 6 semitones (1 semitone = 1/12 octave or about 6% difference between two tones).

When listening to the recordings it appeared that part of the speakers had produced the isolated vowels on a fairly flat tone. Of these tokens the fundamental frequency was not analysed. The other tokens which had a fundamental frequency with a rise-fall pattern were stylised by measuring the fundamental frequency at five points in time and then connecting these points with straight lines. The points of time concerned were the beginning of the rise (t_0), beginning and end of the time interval during which the fundamental frequency is fairly constantly at its top level (t_1 and t_2) "high plateau", end of the fall (t_3) and end of the pitch contour (t_4); between t_3 and t_4 there are no visible pitch changes ("low plateau"). t_0 and t_4 were determined on the analogy of the beginning of the rise time and the end of the decay time respectively in the parallel intensity trace. The time intervals between t_0 and t_1, t_1 and t_2, t_2 and t_3, and t_3 and t_4 were also measured (see also Figure 2).

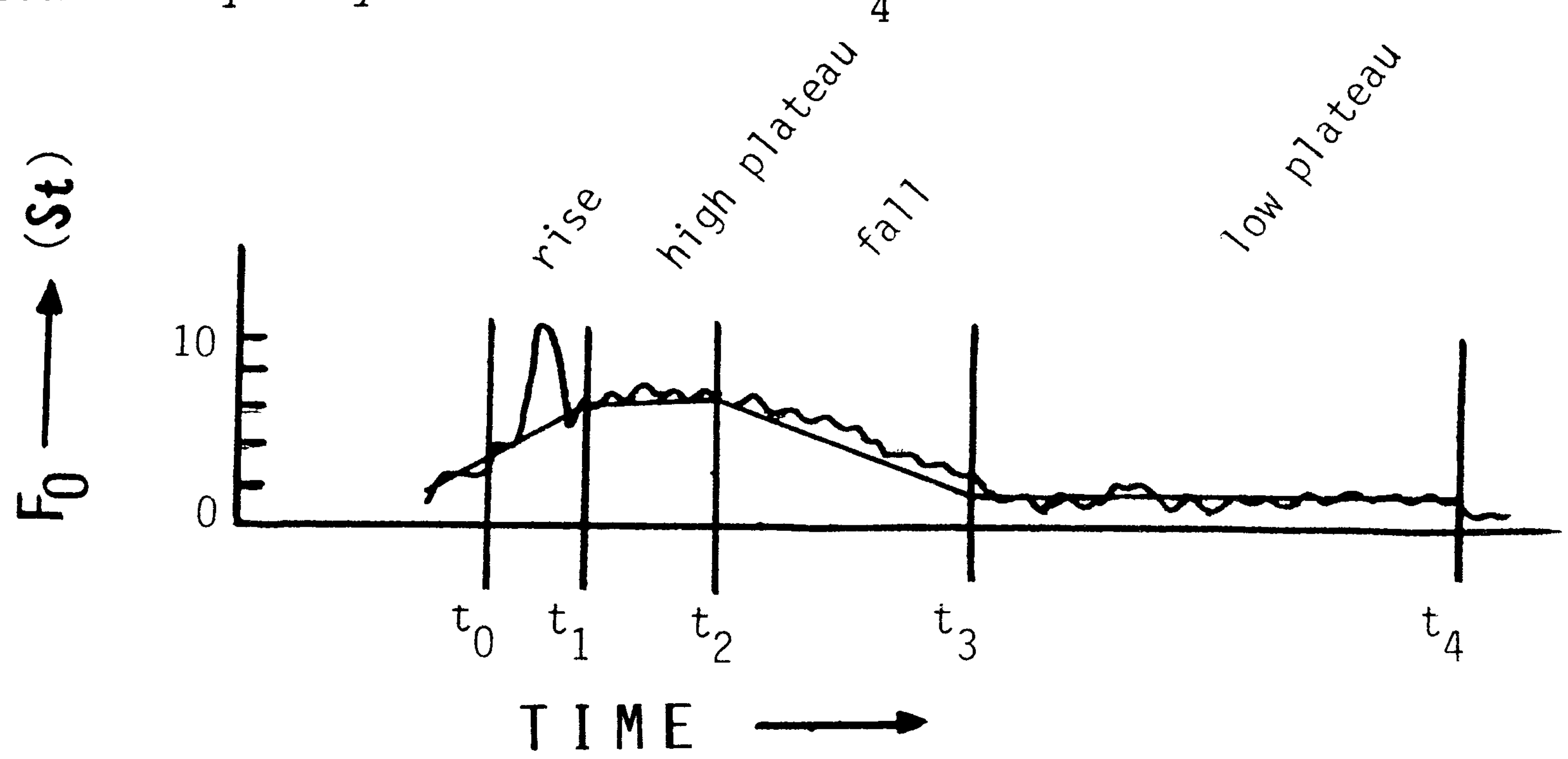

Figure 2: an example of the fundamental frequency contour of a vowel spoken in isolation with a rise-fall pattern and its stylization. Beginning and end of the pitch contour are determined on the analogy of the beginning and the end of the parallel intensity trace (Fig. 1).

3. RESULTS OF THE MEASUREMENTS

3.1. FORMANT STRUCTURE

a. *Vowels in Isolation*

Table 1 gives the means ($\bar{x}$) and standard deviations (s) of the measured formant values of the vowels spoken in isolation (ten speakers, one utterance per speaker). For a breakdown per speaker see Appendix 2.a.

	F_1		F_2		F_3	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
/i/	287	50	2413	248	2860	365
/e/	459	139	1730	341	2435	266
/a/	877	92	1529	126	2447	297
/o/	471	137	1015	208	2263	485
/u/	320	75	914	197	2406	261
/ə/	486	87	1493	112	2647	195

Table 1: mean formant values and standard deviations (s) in Hz of the six vowels spoken in isolation; ten speakers, one utterance per speaker[2].

b. *Vowels Embedded in Monosyllabic Words*

Table 2 lists the measured formant values of the six vowels embedded in monosyllabic words, averaged first over the five utterances per speaker and then over the ten speakers; the standard deviation gives the scatter of the ten speaker means. See also Appendix 2.b.

	F_1		F_2		F_3	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
/i/	361	33	2038	148	2804	161
/e/	611	49	1716	76	2436	138
/a/	848	105	1361	124	2485	161
/o/	587	27	1017	56	2327	221
/u/	375	42	895	63	2571	233
/ə/	416	40	1517	131	2657	148

Table 2: mean formant frequencies ($\bar{x}$, in Hz) and standard deviations (s) of the six vowels embedded in monosyllabic words; ten speakers, five utterances per speaker. The standard deviation (s) gives the scatter of the ten means over the speakers.

3.2. VOWEL DURATION

Table 3 specifies the duration of the vowels in context (except /ə/) averaged first over the five utterances and then over the ten speakers. The standard deviation represents the scatter of the ten speaker means. As for the vowels spoken in isolation (one token per speaker), the ten utterances were averaged. The vowel durations per speaker, averaged over the utterances are also included in Appendix 2.

	in context		in isolation							
	overall duration		overall duration		rise time		steady time		decay time	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
/i/	60	16	335	63	26	16	155	41	154	43
/e/	117	30	340	74	27	16	145	49	168	55
/a/	90	24	344	80	27	23	136	41	181	59
/o/	101	21	366	62	27	16	163	51	176	39
/u/	77	15	339	67	22	11	146	56	171	55
/ə/	-	-	357	70	29	31	162	60	166	53

Table 3: mean duration values and standard deviations (in ms) of the six vowels. The durations of the vowels in context are averaged first over the five utterances and then over the ten speakers. The durations of the vowels spoken in isolation (one token per speaker) are averaged over the ten speakers. The standard deviation (s) expresses the scatter of the means over the speakers.

3.3. FUNDAMENTAL FREQUENCY

The pitch contour of a vowel was only analysed if its fundamental frequency showed a clear rise-fall pattern. Such a stylised pitch contour could be determined completely, or almost completely for the vowels spoken in isolation by three speakers; in these cases the fundamental frequency was measured at the points of time t_0, t_1, t_2, t_3 and t_4, as were the time intervals between these points (see also Figure 2). The data of two of the other speakers could only partially be measured. Five speakers produced the isolated vowels on a fairly flat tone. Here the F_0 was not analysed. The results of the measurements are included in Appendix 2.c.

4. DISCUSSION OF THE RESULTS OF THE MEASUREMENTS

4.1. FORMANT STRUCTURE

Figure 3 contains the six vowels in context with the realisations of each speaker plotted in a F_1/F_2-plane with logarithmic axes. A visual impression of the scatter of each vowel type has been given by connecting the most extreme tokens of each type by a minimal number of solid straight lines, such that we get a demarcation without indentations of the area within the F_1/F_2-plane which contains all the tokens of the same vowel type (cf. Ladefoged, 1967).

"Centres of gravity" (i.e. the points which correspond with the intersections of the mean F_1 and F_2 values of all the vowels contained by a polygon) are indicated with the symbol of the phoneme concerned. The polygons of the vowels spoken in context are shaded. In addition, the scatter of the six vowels spoken in isolation is specified in Figure 3. The extreme tokens per vowel are connected by dashed lines.

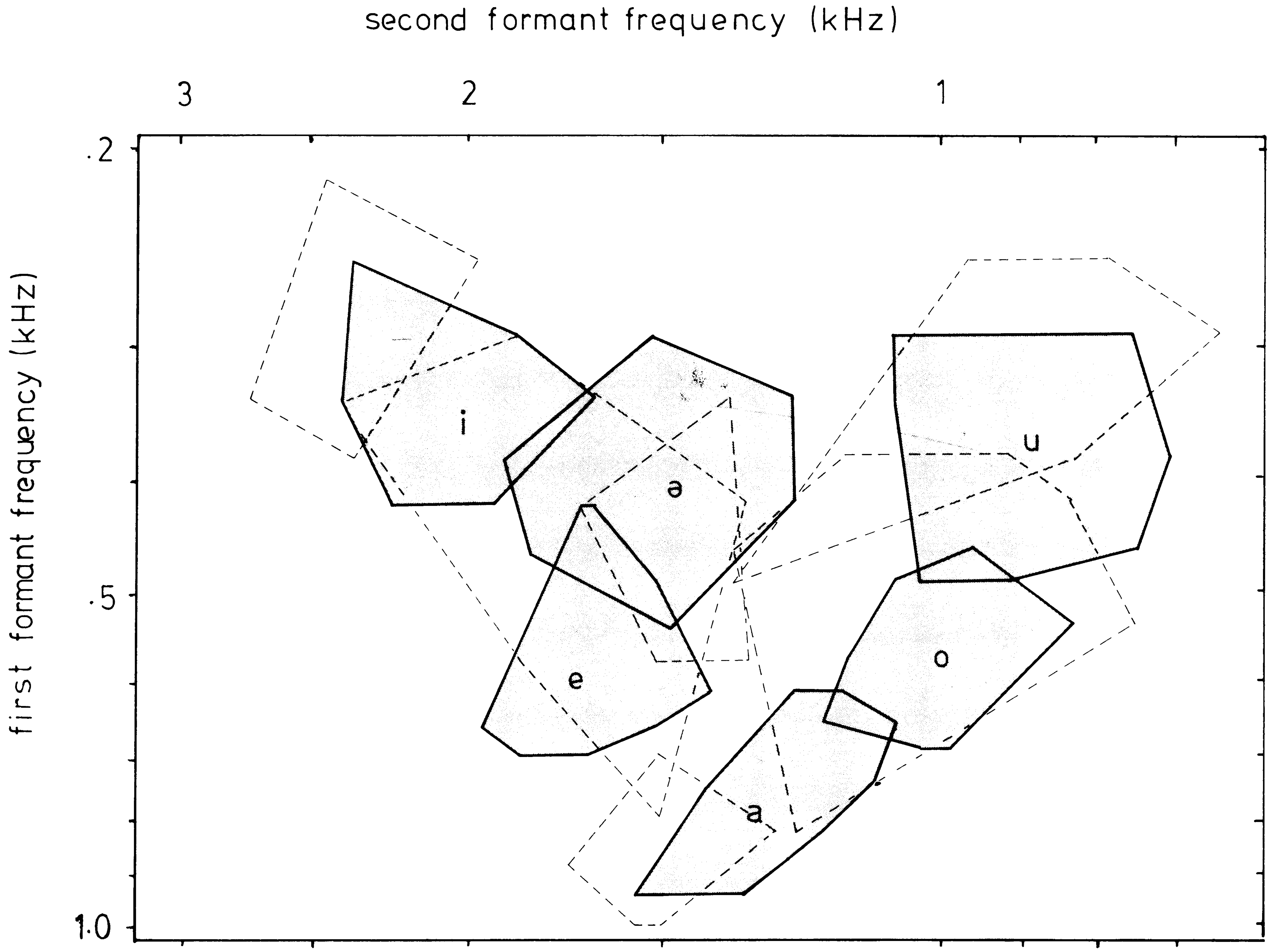

Figure 3: scatter diagram of the six vowels spoken in context and in isolation. The shaded polygons with their continuous straight demarcation lines form the envelopes of the fifty realisations per vowel in context (ten speakers, five realisations per speaker). The centres of gravity (i.e. the intersections of the F_1 and F_2 values of the vowels concerned) are indicated with the symbol of the phoneme concerned; F_1 and F_2 values are averaged first over utterances per speaker and then over the speakers. The polygons with dashed demarcation lines envelop the ten tokens of each of the vowels spoken in isolation (ten speakers, one token per speaker).

The graph allows a number of observations:

1. There is hardly any overlap between the realisations of the six vowels spoken in context, even when all realisations per vowel type are taken into consideration (N = 300).
2. The considerable scatter of /e/, /o/ and /u/ spoken in isolation and the resulting greater overlap of vowels in isolation is remarkable; especially /ə/ spoken in isolation largely falls within the scatter area of one of the other vowels, viz. /e/.
3. It is not possible to simply ascribe the difference between vowels spoken in context and vowels in isolation to reduction phenomena. As regards the corner-vowels /i/, /u/ and /a/ it can be maintained that the polygons of the vowels spoken in context have generally shifted to the centre of the F_1/F_2-plane when compared to the isolated vowels; but this is certainly not true for /ə/. The "ideal position" of /ə/ (i.e. spoken in isolation) seems to be fairly centralized, but in real speech /ə/ appears as a more closed vowel: when spoken in context it is practically in line with the closed vowels /i/ and /u/.

4.2. VOWEL DURATION

Many languages have intrinsic vowel duration differences which are related to the degree of mouth opening. When all other variables are kept constant the rule obtains that the larger the degree of opening - and consequently the articulatory movement - the longer the vowel. According to Lehiste (1970, 18-19) this may well be a universal linguistic phenomenon. Consequently, in Indonesian, we should expect /a/ to have the longest, /e/ and /o/ an intermediate and /i/ and /u/ the shortest duration.

The means over the speakers of our measurements of durations of vowels in context indeed reveal the expected relatively short durations for the closed vowels /i/ and /u/, and a slightly longer duration for the open vowel /a/. The mid-vowels /e/ and /o/, however, are longer than predicted; they are clearly longer than the open vowel /a/; see Table 3.

For an explanation for this unexpected pattern one should first of all consider that the consonantal context was not identical for all the vowels. The relatively long duration of /o/ can be partly attributed to its consonantal context; Nooteboom and Cohen (1976 , 104/5) report that Dutch /ɔ/ is approximately 10 ms shorter when followed by /k/ than when followed by /t/. If this effect of consonantal context on vowel duration is a universal phenomenon, we may assume that Indonesian /o/, when followed by /k/ in the stimulus word instead of by /t/ as it was in the experiment, would have reached a mean duration value of approximately 90 ms. For the relatively long duration of /e/ in context there is a similar explanation: /e/ is followed by /s/ in the stimulus word, and voiceless fricatives do not shorten the preceding vowel to the extent that voiceless plosives do; Nooteboom and Cohen give a difference of around 10 ms for /I/ or /e:/. Therefore, Table 4 also presents estimated vowel durations after correction for the influence of the final consonant.

	(a)	(b)
/i/	60	60
/u/	77	77
/o/	101	91
/e/	117	107
/a/	90	90

Table 4: (a) mean durations of vowels in context (as Table 3);
(b) as (a) but corrected for final consonant, according to Nooteboom and Cohen (1976)

The differences in duration which we found between vowels in context are very small and it is only by means of a rough division into two categories that they can be related to the degree of opening of the mouth: the closed vowels are as a category approximately 30 ms shorter than the mid-vowels and the open vowels.

For the vowels in isolation, however, there is no relation at all between duration and degree of opening of the mouth. None of the differences concerned (see Table 3 or Figure 4 (open circles)) is statistically significant. The fact that there is no relation between vowel duration and degree of opening of the mouth is not caused by the production in isolation: Govaerts (1974, Table 1.15) did find a positive relation between degree of mouth opening and vowel duration for Southern Dutch vowels spoken in isolation (see Figure 4 (plusses)).

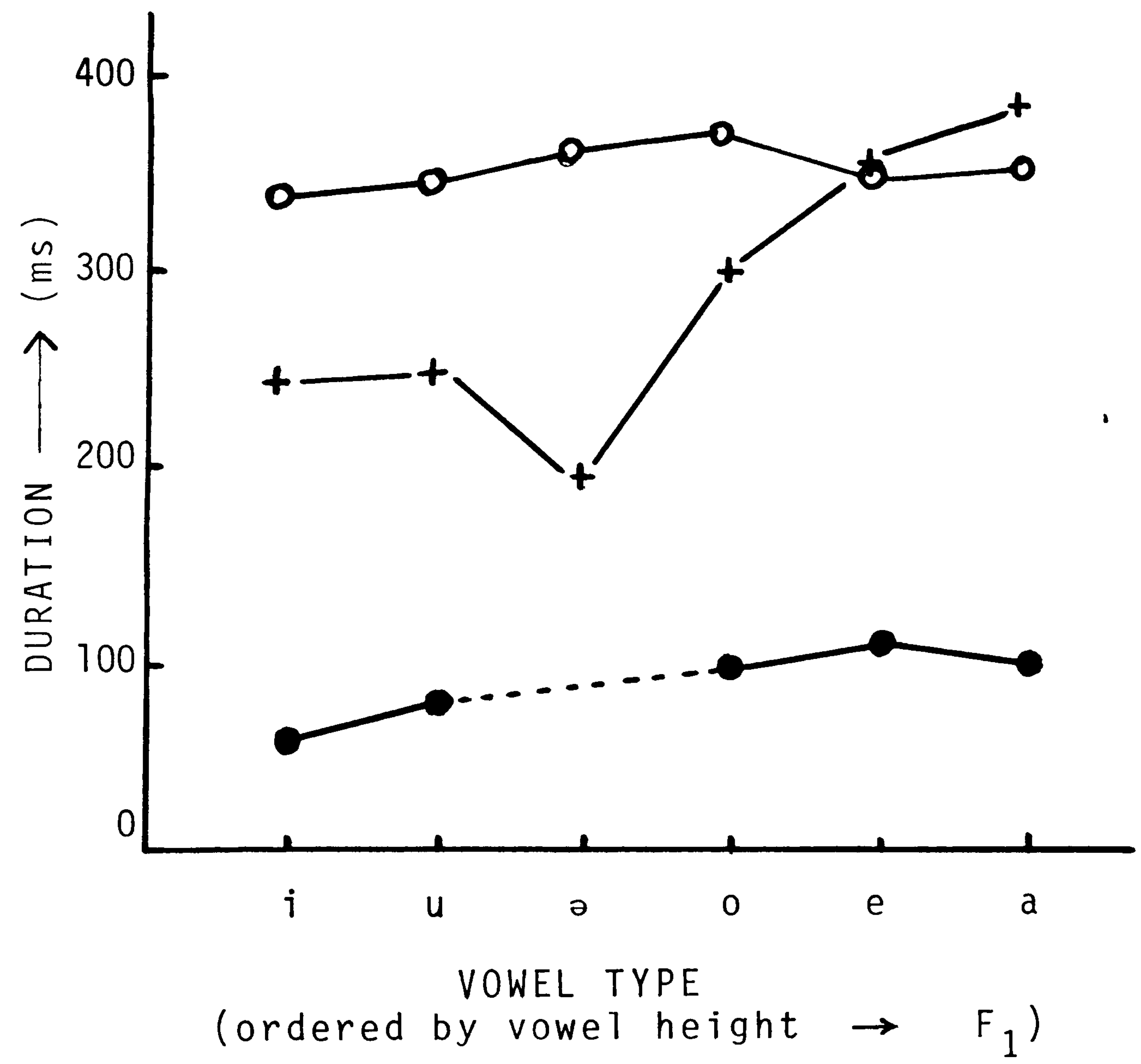

Figure 4: mean duration values of the six vowels in context after correction for the influence of the final consonant (●) and spoken in isolation (O) as a function of degree of mouth opening. For the sake of comparison, duration values of Southern Dutch vowels spoken in isolation (+) have been included (adapted from Govaerts, 1974).

4.3. FUNDAMENTAL FREQUENCY

The data on the fundamental frequency contours with a rise-fall pattern are scant; see Appendix 2.c. For the three speakers for whom such a contour could completely or almost completely be measured, the following obtains (see also Figure 2):

a. At the moment of time t_0 there are fairly large differences in pitch between speakers as well as among vowels; the mean value over the three speakers and the six vowels is 3 semitones above 100 Hz.[3]

b. The mean pitch at the "high plateau" of the contour, i.e. between t_1' and t_2, is 7 semitones above 100 Hz.

c. Each of the three speakers has a fixed pitch at the "low plateau" of the contour (between t_3 and t_4), which varies between 1.5 and 2.5 semitones above 100 Hz for the different speakers.

For four speakers we can compare the high plateaux of the F_0-contour of the various vowel realisations.

speaker nr.	/i/	/u/	/ə/	/e/	/o/	/a/
3	8.5	9.5	8.5	7.5	6.5	6
5	12	13	11	11.5	9.5	9
8	6.5	6	7	5	5	5
10	9.5	8	8	9	7	7

Table 5: "high plateaux" of the stylized F_0-contours (in semitones above 100 Hz) of the realisations of the six vowels spoken in isolation by the four speakers who used a rise-fall pattern.

Table 5 shows that for these four speakers the F_0 of the closed vowels /i/ and /u/ always reaches a higher top-level than the F_0 of the open vowel /a/ by the same speaker. The difference in pitch between /i/ and /u/ on the one hand and /a/ on the other varies for the same speaker from 1 to 4 semitones. Such a variation in fundamental frequency related to the degree of mouth opening is well known in the literature; see, for instance, Lehiste (1970, 68/9), who also claims that differences of sizes as found in Table 5 are without any doubt perceptible.

Mid vowels have a F_0 with a position between the F_0's of open and closed vowels according to Lehiste. In our experiment, however, we find that the F_0 of /o/ agrees with the F_0 of /a/ of the same speaker. The position of /e/ is not clear in this respect.

The F_0 of /ə/ behaves in the same way as the F_0 of the closed vowels /i/ and /u/; the "high plateau" of its contour is approximately as high as that of /i/ and /u/ of the same speaker; it is always higher than the high plateau of /o/ - and definitely higher than that of /a/.

5. EVALUATION OF THE METHOD OF MEASURING

In this explorative study the six Indonesian monophthongs were analysed as spoken by a not randomized and - when one takes into consideration the great diversity of speakers of the Indonesian language - limited number of Indonesian speakers. The ten speakers, all male, were asked to pronounce the vowels embedded in words which again were - in pre-pausal position - inserted in a carrier-sentence, and thereafter in isolation, once again in pre-pausal position in a carrier-sentence. The purpose of using carrier-sentences was to keep the prosody as identical as possible. However, in spite of our precautions, a deviating, rising intonation contour was occasionally found at the end of the sentence.

The vowels in context were only measured in mono-syllabic stimulus words. The form of the stimuli, viz. CVC, where C is a voiceless plosive, sometimes caused a considerable reduction both in vowel duration and in formant values. Occasionally we had to resort to a comparison with section spectrograms of other realisations of the same vowel by the same speaker in order to be able to measure the formants of a particular vowel.

In some cases it was not possible to measure a formant. Notably F_3 of /u/ in about two-thirds of the cases was too weak to be read from the spectrogram. For the numbers of measurements of the formants per vowel and per speaker see Appendix 2.

The results of the measurements on vowels in isolation should be looked at with some reservation: they are based on one realisation per vowel per speaker only.

6. SUMMARY OF RESULTS

The results give us in a tentative way some idea of the Indonesian vowel system as far as the six monophthongs are concerned. Below the most important results of the experiment are summarized.

1. In the F_1/F_2-plane there is hardly any overlap between the realisations of the six vowels in context, even when all realisations are taken into consideration.

2. The "pepet" /ə/, which is generally regarded as a central vowel in Indonesian textbooks, does comply with this qualification when spoken in isolation. When spoken in context, however, /ə/ appears to be a more closed vowel, with an F_1 which is almost as low as the F_1 of the closed vowels /i/ and /u/.

3. From the measurements on the fundamental frequency - which were for that matter obtained for 4 speakers only - the F_0 of /ə/ seems to agree with the F_0 of /i/ and /u/. If /ə/ is indeed to be counted among the closed vowels on account of its low F_1, this would fit in nicely with the data on its fundamental frequency.

REFERENCES

Debrock, M. 1977. "An Acoustic Correlate of the Force of articulation", *Journal of Phonetics*, 5:61-80.

Govaerts, G. 1974. *Psychologische en Fysische Strukturen van Perceptueel Geselekteerde Klinkers; een Onderzoek aan de Hand van Zuidnederlandse Klinkers.* Diss. K.U. Leuven.

Halim, Amran, 1974. *Intonation in Relation to Syntax in Bahasa Indonesia.* Jakarta: Djambatan.

Ladefoged, P. 1967. "The Nature of Vowel Quality: 3. The Acoustic Analysis of Cardinal Vowels," in P. Ladefoged: *Three Areas of Experimental Phonetics*, London: Oxford University Press, 75-103.

———. 1975. *A Course in Phonetics.* New York: Harcourt, Brace and Jovanovich.

Lehiste, I. 1970. *Suprasegmentals.* Cambridge (Mass.)-London: The M.I.T. Press.

Macdonald, R. Ross and Soenjono Dardjowidjojo. 1967. *A Student's Reference-grammar of Modern Formal Indonesian.* Washington: Georgetown University Press.

Nooteboom, S.G. and A.Cohen. 1976. *Spreken en Verstaan; een Inleiding tot de Experimentele Fonetiek.* Assen-Amsterdam: Van Gorcum.

Ro'is, J. Muh. Arsath. 1980. *Bahasa Indonesia; Praktisch Leerboek voor Beginners.* Amsterdam-Antwerpen: Keesing.

Teeuw, A. 1978. *Leerboek Bahasa Indonesia.* Groningen: Wolters-Noordhoff.

Zanten, E. van and V.J. van Heuven. 1983. "A cross-dialect study of vowel perception in Standard Indonesian," in A. Cohen and M.P.R. van den Broecke (eds.): *Proceedings of the Tenth International Congress of Phonetic Sciences*, Foris, Dordrecht (to appear).

Zubkova, L.G. 1967. [The spectral characteristics of Indonesian vowels], *Vestnik Leningradskogo Gosudarstvennogo Universiteta (LGU)*, no. 2.

APPENDIX 1: THE STIMULUS MATERIAL

In the carrier sentence "*Ini kata ...*", 'This is the word ...', the following stimulus words were presented to the speakers:

tik	/tik/	'tap', 'type'
titik	/titik/	'dot', 'drop'
pes	/pes/	'plague'
tetes	/tetes/	'drop'
pak	/pak/	'packet'
kakak	/kakak/	'older sister, brother'
pot	/pot/	'(flower)pot'
pokok	/pokok/	'essence'
kuk	/kuk/	'yoke'
tutup	/tutup/	'closed'
ke	/kə/	'towards'
Tebet	/təbət/	'quarter of Jakarta'

The six monophthongs /i, e, a, o, u, ə/ were presented in the sentence frame "*Dalam kata* ... (*tik, pes, ...*) *terdapat bunyi* ... (*i, e,...*)", 'In the word ... (*tik, pes, ...*) we find the sound ... (i, e, ...)'.

APPENDIX 2

a. *formant values in Hz and vowel durations in ms per speaker of the six vowels spoken in isolation*

vowel	*nr. of speaker*	F_1	F_2	F_3	*overall duration*	*rise time*	*steady time*	*decay time*
/i/	1	377	2345	2837	330	10	180	140
	2	295			400	10	240	150
	3	254	1976	2591	270	20	130	120
	4	254		2345	360	40	130	190
	5	336	2714	3206	320	20	130	170
	6	336		2591	210	20	100	90
	7	213	2427		430	10	200	220
	8	254			350	50	140	160
	9	254	2468	3206	310	50	160	100
	10	295	2550	3247	370	30	140	200
/e/	1	336	2386	2837	350	20	200	130
	2	787	1525	2140	390	10	240	140
	3	377		2673	250	20	110	120
	4	418	1320	2427	360	20	110	230
	5	295	1853	2427	320	30	110	180
	6	459	1730	2386	230	50	90	90
	7	582	1853	2837	480	20	190	270
	8	459		2140	280	10	130	140
	9	459	1812	2304	390	60	150	180
	10	418	1361	2181	350	30	120	200

vowel	*nr. of speaker*	F_1	F_2	F_3	*overall duration*	*rise time*	*steady time.*	*decay time*
/a/	1	992	1525	2386	310	10	150	150
	2	951	1607	2304	340	10	90	240
	3	828	1648	2386	300	40	90	170
	4	992	1566	2263	460	60	200	200
	5	787	1402	2304	360	10	160	190
	6	869	1730	3165	200	10	90	100
	7	910	1484	2304	450	10	180	260
	8	705	1525		300	20	100	180
	9	828	1279	2468	300	70	150	80
	10	910	1525		420	30	150	240
/o/	1	828	1238		390	10	240	140
	2	377	1074	1812	390	20	240	130
	3	459	1361		310	40	100	170
	4	377	1156		450	50	180	220
	5	418	828	1894	310	30	120	160
	6	500	992	2796	260	30	100	130
	7	377	910		420	10	180	230
	8	418			320	10	130	180
	9	541	746	2550	390	50	170	170
	10	418	828		420	20	170	230
/u/	1	500	1361		330	20	200	110
	2	336	828		440	10	260	170
	3	295	828	2591	250	30	80	140
	4	254	787	2222	390	40	100	250
	5	377	828		300	20	110	170
	6	336	1115		280	20	100	160
	7	254	951		420	10	180	230
	8	295	828		300	10	170	120
	9	295	664		280	40	130	110
	10	254	951		400	20	130	250
/ə/	1	582	1443	2591	350	20	170	160
	2	582	1320	2509	450	20	310	120
	3	459	1484	2796	300	20	150	130
	4	582	1525	2345	330	20	130	180
	5	336	1361	2796	310	10	140	160
	6	418	1689	2878	260	30	100	130
	7	500	1566	2468	480	20	170	290
	8	418	1484	2796	360	10	160	190
	9				370	110	130	130
	10	500	1566		370	20	140	210

APPENDIX 2

b. *formant values (in Hz) and vowel durations (in ms) of the six vowels embedded in monosyllabic words, averged over the five utterances per speaker*

		F_1			F_2			F_3			*duration*		
vowel	*nr. of speaker*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*
/i/	1	377	29	5	2042	62	5	2624	45	5	77	6	5
	2	398	41	4	1976	58	4	2735	24	4	45	6	4
	3	336	0	5	2058	41	3	2722	34	5	46	7	5
	4	328	18	5	1755	99	5	2517	128	5	47	8	5
	5	352	22	5	2312	73	5	2903	135	5	83	13	5
	6	385	34	5	2025	168	5	2804	61	5	47	7	5
	7	295	29	5	2189	128	5	3073	275	4	50	6	5
	8	393	22	5	1960	62	5	2845	53	5	59	6	5
	9	361	22	5	1960	74	5	2927	73	5	72	5	5
	10	385	34	5	2099	123	5	2894	111	5	71	13	5

		F_1			F_2			F_3			*duration*		
vowel	*nr. of speaker*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*	$\bar{x}$	*s*	*valid cases*
/e/	1	639	37	5	1820	45	5	2419	45	5	121	6	5
	2	633	21	4	1679	97	4	2550	58	4	84	10	4
	3	623	29	5	1714	103	5	2763	98	5	99	12	5
	4	648	37	5	1673	173	5	2288	90	5	155	20	5
	5	672	34	5	1845	34	5	2296	61	5	118	20	5
	6	500	116	4	1710	191	4	2407	71	4	83	12	4
	7	582	50	5	1681	61	5	2458	202	4	81	4	5
	8	639	47	5	1574	73	5	2394	124	5	155	12	5
	9	574	18	5	1730	50	5	2353	53	5	113	9	5
	10	598	22	5	1730	145	5	2435	98	5	145	9	5
/a/	1	894	22	5	1451	18	5	2304	50	5	99	6	5
	2	910	41	3	1293	24	3	2810	24	3	76	15	4
	3	795	53	5	1246	89	5	2566	273	5	82	6	5
	4	1000	67	5	1580	103	3	2478	178	4	56	12	5
	5	779	45	5	1377	85	5	2288	69	5	138	13	5
	6	877	67	5	1443	92	5	2575	160	5	75	9	5
	7	918	79	5	1336	22	5	2288	144	5	74	11	5
	8	689	47	5	1279	41	5	2484	111	5	96	21	5
	9	689	47	5	1156	65	5	2517	128	5	105	7	5
	10	926	74	5	1451	61	5	2542	146	5	102	6	5
/o/	1	582	58	5	1115	96	5	2355	52	4	104	14	5
	2	551	91	4	982	39	4	1966	283	4	80	7	4
	3	574	67	5	976	37	5	2171	78	4	91	4	5
	4	607	94	5	984	53	5	2566	210	5	100	12	5
	5	639	37	5	1058	62	5	2386	172	5	131	16	5
	6	574	61	5	1082	67	5	2724	62	4	70	6	5
	7	574	34	5	982	52	4	2230	89	5	84	9	5
	8	566	22	5	1008	22	5	2238	22	5	110	10	5
	9	582	29	5	934	94	5			0	119	16	5
	10	623	58	5	1049	62	5	2304	209	5	122	9	5
/u/	1	328	18	5	943	34	5	2427	58	2	83	4	5
	2	398	24	4	941	91	4	2509		1	61	2	4
	3	344	34	5	828	65	5	2673		1	72	3	5
	4	344	45	5	935	62	5	2325	87	2	75	13	5
	5	402	37	5	836	18	5			0	101	12	5
	6	402	74	5	976	55	5	3083		1	68	23	5
	7	311	22	5	935	158	5	2591		1	68	5	5
	8	451	34	5	828	65	5	2427		1	71	10	5
	9	377	50	5	803	85	5	2530	53	4	83	13	5
	10	393	37	5	926	47	5			0	92	8	5
/ə/	1	492	45	5	1517	67	5	2484	111	5			0
	2	428	21	4	1546	85	4	2550	108	3			0
	3	410	34	5	1599	79	5	2607	103	5			0
	4	361	22	5	1738	106	5	2927	326	5			0
	5	361	55	5	1517	18	5	2607	80	5			0
	6	426	18	5	1615	34	5	2689	69	5			0
	7	410	18	5	1369	45	5	2517	168	5			0
	8	434	37	5	1443	41	5	2845	202	5			0
	9	385	34	5	1271	34	5	2566	37	5			0
	10	451	34	5	1558	155	5	2776	71	4			0

c. *fundamental frequency of the vowels spoken in isolation with a clear rise-fall pattern*

	frequency in semitones above 100 Hz			time in ms			
	beginning rise t_0	high plateau $t_1 - t_2$	low plateau $t_3 - t_4$	rise $t_0 - t_1$	high plateau $t_1 - t_2$	fall $t_2 - t_3$	low plateau $t_3 - t_4$
subj. 3 /i/		8.5	2	30	50	120	60
/e/	3.5	7.5	2	30	40	80	110
/a/	3	6	2	50	20	80	140
/o/		6.5	2				80
/u/	5.5	9.5	1.5	30	40	100	80
/ə/	5	8.5	2	30	60	90	120
subj. 4 /u/	7	10	2.5	70	50	100	140
/ə/		5	2	90		90	150
subj. 5 /i/		12					
/e/		11.5					
/a/		9					
/o/		9.5					
/u/		13					
/ə/		11					
subj. 8 /i/	2	6.5	2	50	50	110	110
/e/		5	1.5	40	70	80	70
/a/	2.5	5	2	50	30	90	90
/o/		5	1.5	50	70	90	70
/u/	2.5	6	2.5	40	50	70	140
/ə/	2	7	2	60	50	100	110
subj. 10 /i/		9.5	2			100	110
/e/	3	9	2	40	70	110	100
/a/	2.5	7	2.5	50	60	90	190
/o/	4	7	2	40	60	110	190
/u/	5.5	8	2	40	70	100	150
/ə/	4	8	1.5	40	70	100	130

NOTES

1. The research reported here was financed in part by the Netherlands Organisation for the Advancement of Pure Research (Z.W.O. project nr. 17-21-20).

2. For the rather variable numbers of tokens underlying the means, see Appendix 2.a.

3. 100 Hs is an arbitrary reference value relative to which all periodicity measurements have been expressed.