VOWEL CONTRAST REDUCTION IN JAPANESE COMPARED TO DUTCH

by Tjeerd de Graaf and Florina J. Koopmans-van Beinum

1. INTRODUCTION

In a previous study on vowel contrast reduction (Koopmans-van Beinum, 1980) a measure was introduced to indicate the degree of acoustic contrast of whole vowel systems in various speech conditions. Application of this measure ASC (Acoustic System Contrast) to the Dutch language revealed that the degree of contrast is to a great extent speech-situation and speaker dependent. The value of ASC decreases when passing from vowels in isolation through different speech situations to unstressed vowels in free conversation. This leads us to the question whether it would be possible to use this measure for the comparison of speech situations and speakers of other languages as well, even when these languages have a numerically different vowel system, or when the individual vowels in the formant field take deviant positions.
We might hypothesize a rule according to which vowel systems involving fewer vowels would have a larger degree of vowel contrast reduction expressed in the parameter ASC than richer vowel systems.
To answer these questions we here compare the degrees of acoustic system contrast (ASC) of speakers of two quite distinct languages, Dutch and Japanese, with vowel systems of respectively twelve and five vowels.

2. DATA COLLECTING

2.1 Dutch Speech Material

The speech material for Dutch was derived from Koopmans-van Beinum (1980). In the present study we will make use of the data of the two male speakers in three speech conditions only for the purpose of comparison with the Japanese data. So we shall use vowels pronounced in isolation, vowels in monosyllabic words, and vowels in stressed syllables from free conversation. One of the two speakers may be considered to be a 'trained speaker', which means that he uses his voice and his pronunciation as part of his profession, whereas for the other speaker a careful pronunciation and phonation does not play a role in his job.

The Dutch vowel system consists of twelve more or less monophthongal vowels and three diphthongs. Although the diphthongs certainly play some role in the whole Dutch vowel system, they are left out of consideration here.

So the twelve vowels of the Dutch vowel system, together with key words are:

[u] as in	'boek'	(book)	[ø] as in	'beuk'	(beech)
[o] as in	'boot'	(boat)	[œ] as in	'buk'	(stoop)
[ɔ] as in	'bod'	(bid)	[i] as in	'biet'	(beet)
[ɑ] as in	'bad'	(bath)	[I] as in	'bit'	(bit)
[a] as in	'baat'	(profit)	[e] as in	'beet'	(bite)
[y] as in	'buut'	(butt)	[ɛ] as in	'bet'	(moisten)

For both Dutch speakers the speech material in this study consists of:

(1) the twelve vowels spoken in isolation (3 series);
(2) the vowels in monosyllabic isolated words (5 series);
(3) the vowels in stressed syllables from free conversation (10 series).

Formant frequencies (F1 and F2) were measured by spectral analysis (Wempe, 1979) at one place in the vowel where the intensity was maximal. Further data on measuring methods and results are given extensively in Koopmans-van Beinum (1980), so here we will confine ourselves to the mean formant frequencies per vowel per speech condition for both Dutch speakers.

2.2 Japanese Speech Material

The phonemic system of Japanese vowels is rather simple: Japanese has only five vowels; its vowel system can be given by the following diagram:

```
      i     u

        e      o

           a
```

with horizontally the feature front-back (tongue position) and vertically the feature open-closed. The vowel [u], which in standard Japanese is pronounced with a fairly advanced tongue position and with spread lips cannot be regarded as a back vowel; the vowel [o] is the only rounded back vowel in Japanese. Further, according to Takebayashi (1975), Japanese vowels have the following characteristic features which distinguish them from other vowel systems like those of English and of Dutch:
(1) The vowels [i] and [u] are very often devoiced when they occur between voiceless consonants and in word-final position.
(2) In Japanese there is no significant distinction between strongly stressed vowels and weakly stressed vowels, since stress does not play any phonological role. It is stated by Takebayashi that the vowels [i], [e], [a], [o] and [u] are always pronounced without serious qualitative distortion, even when they are in relatively weak syllables. This last property can be investigated by measuring the acoustic data of Japanese vowels in different speech conditions as has been done for Dutch by Koopmans-van Beinum (1980).
The five vowels of Japanese, together with key words are:

[u] as in 'kukei' (rectangle)
[o] as in 'koku' (unit of volume)
[a] as in 'kaki' (persimmon)
[i] as in 'kiku' (to hear)
[e] as in 'kekkon' (marriage)

The speech of three male native speakers of Japanese was recorded in the Institute of Phonetic Sciences, Groningen State University. This material consists of the following parts for each speaker:

(1) the vowels [a], [i], [u], [e] and [o] spoken in isolation (5 series)

(2) the vowels in isolated words (5 series)

(3) the vowels in free conversation (5 series)

In speech situation (2) the vowels are produced in bisyllabic words, because the Japanese language does not allow consonant clusters and final consonants in syllables. In this way we get the vowels in a position where they are surrounded by consonants, in most cases by voiceless plosives, like in 'kiku'. These words are read from cards in a given speech tempo and registration is made on tape with the aid of a Sennheiser microphone and a REVOX tape recorder. The same is done for the free conversation between Japanese speakers. From this conversation a random choice is made of five items for each vowel per speaker. The resulting formant values are obtained in all three speech conditions by averaging over the five items.
During the recording session no further instructions were given to the speakers.

2.3 Measuring Japanese Speech Material

The acoustical data for the Japanese vowels were measured by using the Digital Speech Analysis System of the Groningen Institute of Phonetic Sciences. The relevant speech segments were registered from the analog tape on the B&K 7502 Digital Event Recorder. This contains the digitalized acoustic signals in a memory of 10240 8-bit words that can be stored on separate disk files. The input sampling frequency was taken at 20 kHz and the spectral components between 0 and 5 kHz in the analog signal are reproduced appropriately when using D/A conversion. With this method we obtain samples of 0.5 sec. duration in which the relevant vowel part of the speech signal can be recognized. This can be done by listening to the reproduced signal for which the beginning and the end position can be changed on the computer terminal (with a precision of 0.5 msec.), and by looking at the oscillogram that is reproduced on the scope. In this way we determined a more or less stable central segment of the vowel and applied the LPC analysis program that provides the values of the first four formant frequencies and the corresponding band widths per time frame of 512 sampling points (25.6 msec.). These frames are shifted from the beginning to the end of the speech segment in steps of 256 sampling points (12.8 msec.) and from the average value of the two first formants we got values of F1 and F2 that were characteristic for the vowel in question.

3. DATA PROCESSING

The results for the measurements of the formant values can be reproduced on a list or on a plot that is prepared with the aid of the VERSAPLOT software packet on the VERSATEC electrostatic printer-plotter. The formant values F1 and F2 for the speaker centroid C were calculated as the average value calculated over all three speech conditions. In accordance with the convention used by Koopmans-van Beinum (1980) the formant values were transferred to the logarithmic expressions $LF_i = 100\,{}^{10}\log F_i$ with i=1 or 2. With these values the Acoustic System Contrast (ASC) was calculated which gives a measure for the mean square of the distances in the F1-F2 plane between the different formant values for the vowels and the centroid according to the formula:

$$ASC = \frac{1}{N} \sum_{j=1}^{N} (\vec{V}_j - \vec{C})^2$$

in which $\vec{V}_j$ is the two-dimensional vector for the vowel j in the

F1-F2 plane and $\vec{C}$ is the vector for the centroid C.
The values of the Acoustic System Contrast parameter ASC for the different speakers and in the three speech conditions were calculated in the computer and listed in Tables I to V together with the mean formant values.

TABLE I Mean formant frequencies and ASC in three speech conditions together with the centroid frequencies for Japanese speaker 1.

	ISOLATED VOWELS (n=5)		VOWELS IN WORDS (n=5)		VOWELS IN CONVERS. (n=5)	
	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)
[a]	765	1138	730	1351	611	1322
[i]	302	2463	355	2265	348	2233
[u]	295	1071	320	1032	443	1201
[e]	517	2066	475	1873	437	1788
[o]	465	794	462	915	464	1025

CENTROID FREQUENCIES : F_1=466 Hz F_2=1502 Hz
ACOUSTIC SYSTEM CONTRAST: ASC (isol.) = 613
ASC (words) = 383
ASC (conv.) = 206

TABLE II Mean formant frequencies and ASC in three speech conditions together with the centroid frequencies for Japanese speaker 2.

	ISOLATED VOWELS (n=5)		VOWELS IN WORDS (n=5)		VOWELS IN CONVERS. (n=5)	
	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)
[a]	774	1122	753	1255	676	1221
[i]	294	2311	313	2192	301	2098
[u]	326	1378	375	1428	412	1396
[e]	484	1904	424	1880	427	1628
[o]	505	803	481	884	471	1013

CENTROID FREQUENCIES : F_1=468 Hz F_2=1501 Hz
ACOUSTIC SYSTEM CONTRAST: ASC (isol.) = 501
ASC (words) = 363
ASC (conv.) = 254

TABLE III Mean formant frequencies and ASC in three speech conditions together with the centroid frequencies for Japanese speaker 3.

	ISOLATED VOWELS (n=5)		VOWELS IN WORDS (n=5)		VOWELS IN CONVERS. (n=5)	
	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)
[a]	741	1282	624	1372	605	1356
[i]	270	2422	328	2138	359	2087
[u]	298	1215	329	1284	446	1309
[e]	491	1913	443	1884	467	1767
[o]	491	816	509	876	478	1098

CENTROID FREQUENCIES : F_1=459 Hz F_2=1521 Hz

ACOUSTIC SYSTEM CONTRAST: ASC (isol.) = 540
ASC (words) = 322
ASC (conv.) = 149

TABLE IV Mean formant frequencies and ASC in three speech conditions together with the centroid frequencies for Dutch speaker 1.

	ISOLATED VOWELS (n=3)		VOWELS IN WORDS (n=5)		VOWELS IN CONVERS. (n=10)	
	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)
[u]	317	783	308	809	326	903
[o]	421	861	407	817	429	1079
[ɔ]	600	1030	506	959	461	994
[ɑ]	658	1260	657	1266	609	1238
[a]	692	1398	717	1437	642	1429
[y]	292	1902	296	1861	321	1764
[ø]	425	1715	372	1614	454	1535
[œ]	443	1573	451	1446	405	1464
[i]	293	2503	292	2470	309	2370
[I]	403	2258	385	2156	379	2110
[e]	393	2220	392	2129	441	1948
[ɛ]	653	1965	605	1741	575	1649

CENTROID FREQUENCIES : F_1=454 Hz F_2=1574 Hz

ACOUSTIC SIGNAL CONTRAST: ASC (isol.) = 426
ASC (words) = 418
ASC (conv.) = 272

TABLE V Mean formant frequencies and ASC in three speech conditions together with the centroid frequencies for Dutch speaker 2.

	ISOLATED VOWELS (n=3)		VOWELS IN WORDS (n=5)		VOWELS IN CONVERS. (n=10)	
	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)	F_1(Hz)	F_2(Hz)
[u]	317	773	325	761	336	990
[o]	443	875	454	942	459	1135
[o]	468	907	533	1052	458	1084
[α]	685	1217	638	1281	533	1225
[a]	750	1338	707	1385	572	1226
[y]	298	1760	307	1679	311	1680
[ø]	465	1522	459	1513	457	1421
[œ]	467	1515	438	1490	408	1478
[i]	297	2335	302	2264	315	2063
[I]	378	2143	394	1977	387	1849
[e]	425	2057	430	1928	491	1768
[ε]	538	1897	560	1754	456	1726

CENTROID FREQUENCIES : F_1=453 Hz F_2=1500 Hz
ACOUSTIC SIGNAL CONTRAST: ASC (isol.) = 400
ASC (words) = 310
ASC (conv.) = 178

The Figures 1 to 3 demonstrate the results for the Japanese speakers used in our experiment; these can be compared to Figs. 4 and 5 that are produced with the data of the two male speakers given by Koopmans-van Beinum (1980). In these figures the centroid is indicated by C.

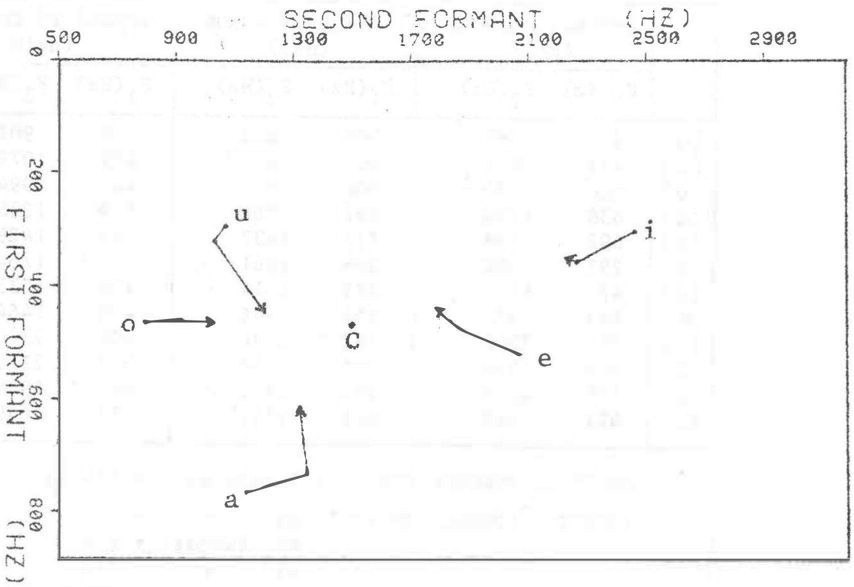

Fig. 1 Japanese speaker 1: mean formant frequencies per vowel per speech condition (connected per vowel from isolation via words to free conversation) in the F_1 - F_2 plane.
C indicates the speaker centroid.

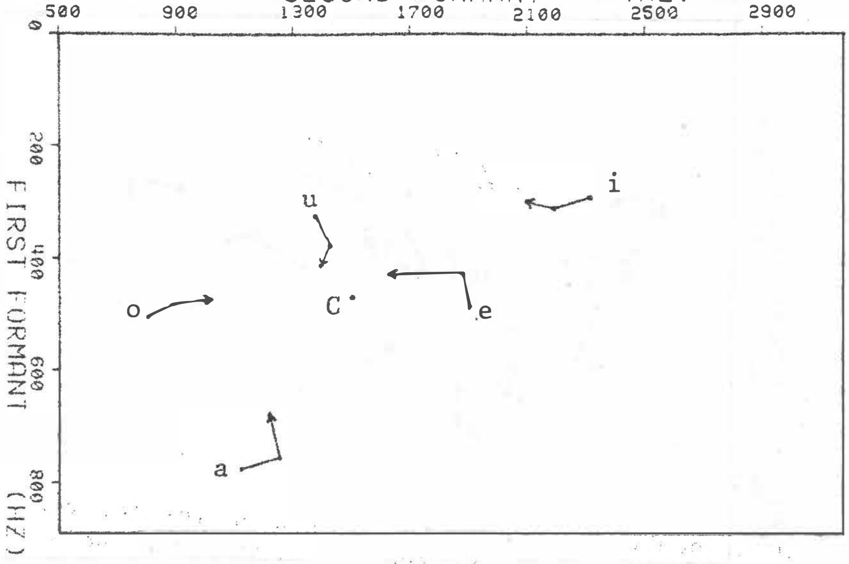

Fig. 2 Japanese speaker 2: mean formant frequencies per vowel per speech condition (connected per vowel from isolation via words to free conversation) in the $F_1 - F_2$ plane. C indicates the speaker centroid.

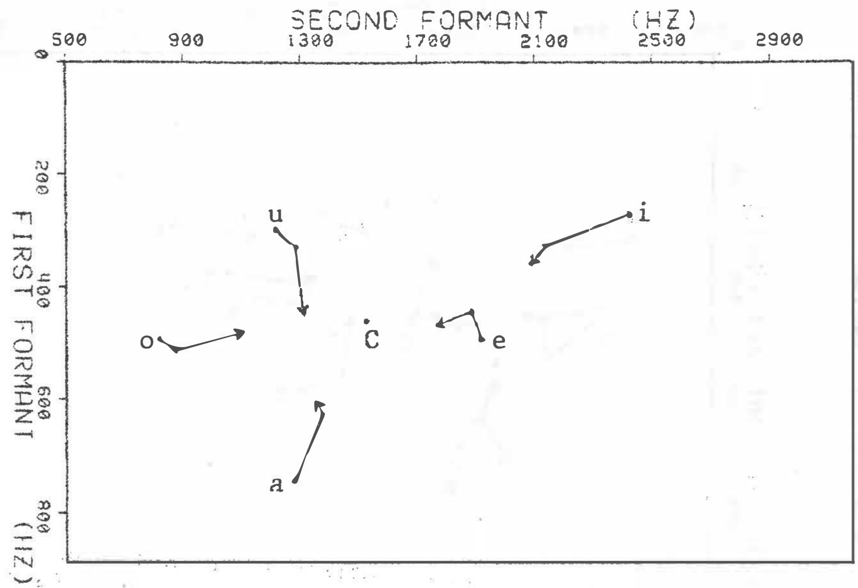

Fig. 3 Japanese speaker 3: mean formant frequencies per vowel per speech condition (connected per vowel from isolation via words to free conversation) in the $F_1 - F_2$ plane. C indicates the speaker centroid.

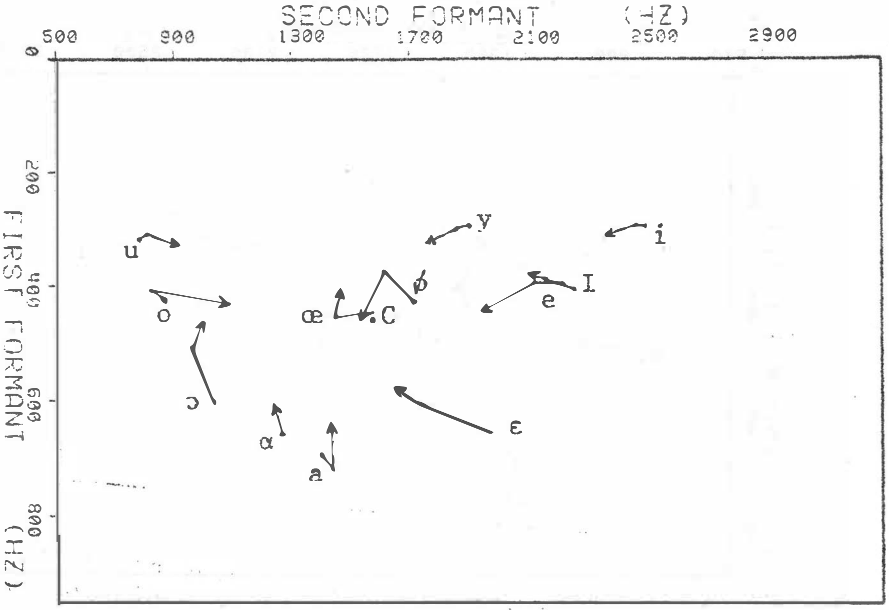

Fig. 4 Dutch speaker 1: mean formant frequencies per vowel per speech condition (connected per vowel from isolation via words to free conversation) in the $F_1 - F_2$ plane. C indicates the speaker centroid.

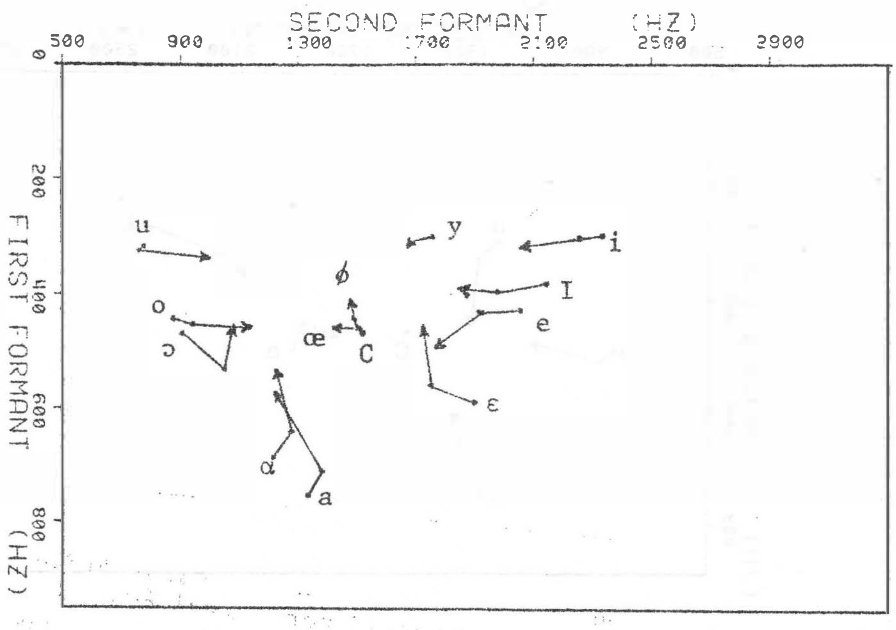

Fig. 5 Dutch speaker 2: mean formant frequencies per vowel per speech condition (connected per vowel from isolation via words to free conversation) in the $F_1 - F_2$ plane. C indicates the speaker centroid.

4. CONCLUSIONS AND DISCUSSIONS

In Japanese as well as in Dutch speech the centroids have formant values F1 and F2 close to the line F2 = 3 x F1 in the F1-F2 plane, where we find the position of the 'neutral vowel' corresponding to the neutral position of the vocal tract. Elsewhere we show that in general this line is still further approached by the centroid if we apply a weighting factor for each vowel, related to the frequency of occurrence in normal speech (Koopmans-van Beinum, 1981, and forthcoming). It appears that this property holds quite generally and is not language dependent.
In the case of Japanese speech the more central position of the vowel [u] compared to Dutch is also shown in the acoustic parameters: the corresponding F2 values are higher in Figs. 1 to 3 than in Figs. 4 and 5.
The ASC values for Japanese decrease considerably from speech situation 1 (isolated vowels) to speech situation 3 (vowels in conversation): for speaker JAP1 by a factor 2.98, for speaker JAP2 by a factor 1.97 and for speaker JAP3 by a factor 3.62. For the two Dutch speakers NED1 and NED2 these values are 1.57 and 2.25, respectively.
From this material we cannot draw the conclusion that for Japanese the vowel contrast reduction is systematically greater than for Dutch and certainly not that a language with five vowels has a greater amount of vowel contrast reduction than a language with twelve or more vowels. We find, however, that a considerable reduction takes place (in contradiction with the statement of Takebayashi). This is probably a universal property of all languages, whereas the quantitative data expressed in parameters like the Acoustic Signal Contrast (ASC) are to a great extent speaker dependent.

REFERENCES

Koopmans-van Beinum, F.J. (1980). Vowel Contrast Reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions. Diss., University of Amsterdam.

Koopmans-van Beinum, F.J. (1981). How can a measure for acoustic system contrast be related to speech production? Proceedings of the Fourth FASE Symposium on Acoustics and Speech, Venezia, Vol. I, E.S.A. Roma, pp. 123-126.

Koopmans-van Beinum, F.J. (forthcoming). Systematics in Vowel Systems. (To be published in a Festschrift in honour of Prof. Dr. A. Cohen.)

Takebayashi, S. (1975). The Vowels of Japanese and English. Lexicon, no. 4, pp. 49-67.

Wempe, A.G. (1979). An experimental segment spectrograph based on some notes on frequency analysis of speech segments. Proceedings of the Institute of Phonetic Sciences, Amsterdam 5, pp. 44-102.