〔RESEARCH〕

A Corpus of Japanese Vowel Formant Patterns

Parham MOKHTARI　Kazuyo TANAKA

This paper describes a dataset of formant patterns measured in the steady-states of recorded Japanese vowels. Five adult, male, native speakers of Japanese were selected from the "ETL-WD-I and II" balanced word dataset; and for each of the five vowels /i, e, a, o, ɯ/, 22 different words were selected on the basis of consistently finding the lengthiest and most steady-state vocalic nuclei. A semi-supervised method based on linear-prediction (LP) analysis of the speech waveform was then used to carefully extract the first four formants in five consecutive frames of each vocalic nucleus, thereby yielding a total of 2750 patterns of formant-frequencies {F1, F2, F3, F4} and formant-bandwidths {B1, B2, B3, B4}. These formant patterns are offered in electronic form, with the aim of contributing to the small but growing body of publicly available formant data.

§1 Introduction

The formants, or acoustical resonances of the vocal-tract, have long been regarded as one of the most compact and descriptively powerful parameter-sets for voiced speech sounds, with important correlates in both the auditory-perceptual and articulatory domains. However, notwithstanding the importance of these acoustic-phonetic parameters, to this day they remain difficult to measure reliably and robustly by automatic methods alone, leaving little choice but for manual or semi-supervised methods which are usually labour-intensive. A carefully-constructed corpus of formant patterns, i.e., ordered sets of formant frequencies {F1, F2, F3, …} and formant bandwidths {B1, B2, B3, …}, ought therefore to be of special value to the speech research community, and should be made publicly available for the advancement of speech science.

Perhaps the most well-known corpus of formant patterns is that reported in Peterson and Barney's[1] (1952) classic study of the first three formant frequencies of 10 vowels recorded by speakers of American English. However, it was only after about four decades that the full body of that famous dataset (as distinct from the speaker-averaged data subsequently used in many studies) was restored and made publicly available in electronic form[2]. More recently, Clermont and Barlow have restored[3] and made available on the World Wide Web[4] (WWW), a corpus of Australian English vowels' formant patterns originally measured by Bernard[5] in 1967 and, similarly to the Peterson & Barney data, hitherto reported only in speaker-group- (or idiolect-) averaged form. Meanwhile, numerous formant datasets have been reported in the literature but, perhaps owing mainly to practical constraints, not readily made publicly available. Distinguished amongst such datasets are those reported by Fant[6] for Swedish vowels, by Fujisaki et al.[7] for Japanese vowels, by Pols and his colleagues[8,9] for Dutch vowels, by Broad et al.[10,11] for 30 steady-state vowels recorded by phonetically-trained speakers, and more recently by Hillenbrand et al.[12] who replicated and extended Peterson and Barney's data.

Naturally, those corpora were each constructed with specific research motivations which determined the number and types of speakers, the phonetic complexity of the speech materials, the number of formants measured, and other experimental conditions. Our own motivations for constructing a new dataset of formants,

KEY WORDS : formants, vowels, Japanese

stem primarily from research aimed at developing a computer text-to-speech synthesis or speech-to-speech conversion system, which should have the flexibility to synthesise speech with a wide range of easily-selectable, personal characteristics. In that research context, it is relevant to note that most of the concatenative speech synthesisers currently in vogue thanks to their ability to produce natural-sounding speech, are heavily data-dependent, implying that speech with a new speaker's characteristics can be synthesised only after processing a range of utterances recorded specifically by that speaker. By contrast, we have proposed[13] an acoustic-articulatory model of inter-speaker variability which, after being trained on the data of a number of speakers, can then be used to simulate a wide range of personal characteristics along principal, articulatory dimensions of that variability.

Indeed, consistent with the notion that the fundamental sources of inter-speaker variability are in the articulatory (or speech-production) domain, the parameters of our model describe that variability in terms of the shape and length of the speakers' vocal-tract (VT) area-functions (the cross-sectional area of the VT airway as a function of the distance from the glottis to the lips). Moreover, as the acquisition of direct articulatory measurements (e.g., by means of magnetic-resonance or ultrasound imaging) is difficult and generally impractical for a speech synthesis or speech conversion system, we prefer that the VT area-functions used in our model are determined by acoustic-to-articulatory mapping (or speech inversion). While we are currently investigating methods of estimating the VT area-function directly from a more easily-measured (e.g., whole-spectrum) representation, a core dataset of reasonably realistic VT area-functions is constructed first using our method of inversion[14] (a hybrid method based on the linear-prediction (LP) model and an extended version of the Schroeder-Mermelstein[15] (SM) parameterisation of VT-shapes), which requires careful measurements of at least the first three or four formant frequencies and bandwidths. In this regard, most previous studies reporting formant data (including the prominent ones cited earlier) are not helpful, as the relative unreliability of bandwidth measurements, and their admittedly secondary role from acoustic-phonetic and auditory-perceptual points of view, have in most cases precluded them from being reported alongside the formant frequencies. However, our method of inversion (i) requires the bandwidths to resolve the well-known problem of nonuniqueness[16], and (ii) relies on a sufficiently rich dataset to provide a statistically reliable estimate of the bandwidths' mean values on a per-vowel basis.

In this paper we therefore report a new corpus of formant frequencies and bandwidths, which firstly meets our immediate research requirements, and may then be made publicly available (via the WWW) in the hope of encouraging similar sharing of useful data amongst speech researchers. In the next section we describe the selection of speech materials and speakers, in section 3 we describe the methods, problems and corrections involved in formant estimation, and in section 4 we conclude with directions for accessing our formant data on the internet.

§ 2 Speech Materials and Speakers

The speech materials were taken from the "ETL-WD-I and II" dataset which comprises a phonetically-balanced list of 1542 words containing VCV/CVC* combinations[17]. Those words were originally recorded at Electrotechnical Laboratory in 1985 and 1987, by 10 adult male, native speakers of Japanese. Furthermore, the acoustic speech waveforms had already been acoustic-phonetically segmented, with a list of the start, the duration, and the phonetic identity of every segment stored in a separate computer file (henceforth referred to as a "segmentation file") for each word of each speaker. For the purposes of our current research aimed at acoustic-articulatory modelling of inter-speaker differences, we specifically sought the most steady-state vocalic nuclei /i, e, a, o, ɯ/ with the least coarticulatory

＊ V and C denote a *vowel* and a *consonant*, respectively.

influences from adjacent segments.

To that end, we first conducted an automated search of the romanised phonetic representations of all 1542 words, in order to identify for each of the five vowels independently, those words which contain either a long-vowel (denoted by "V–"), a double- or repeated-vowel (denoted by "VV"), or a vowel in quasi-neutral preceding context (denoted by "hV"); where in the latter case any instances of "ch" or "sh" were excluded, and in all three cases any instances of potentially nasalised context (i.e., either preceding or following "n", "N", "m" or "g") were excluded. Owing to recording and other problems identified at a later stage of processing, the two words with codes W0494 and W1014, and the two words with codes W1332 and W1524 were excised from the resulting lists for the vowels /a/ and /o/, respectively. Our search then yielded the following numbers of distinct words for each vowel: 98 for /i/, 58 for /e/, 88 for /a/, 277 for /o/, and 156 for /ɯ/. It is interesting to note that in this phonetically-balanced list of Japanese (and Japanised) words, there is a significantly higher number of words which contain long-vowel, double-vowel, or quasi-neutral instances of the vowel /o/ in particular.

Next, for each of the five vowels and each of the 10 speakers, the segmentation file for each word in the list was queried automatically in order to extract the duration (in msec) of the relevant vocalic nucleus, denoted by "VVV" in each segmentation file. The word-list for each vowel was then sorted in decreasing order of the *shortest* vowel-nucleus duration across the 10 speakers, whereupon it was found that in order to secure a duration *no less than 100msec* for any given nucleus, only the top 22 words in each ordered list could be accepted (as dictated by the ordered list for the vowel /a/ in particular, for which durations of 95msec and less were found starting from the 23rd word). Computed across those 22 words in each list, the mean and range of the shortest vowel durations amongst the 10 speakers are shown in **Table 1**. It is interesting to note that the vocalic nucleus of /o/ in particular has the longest of the mean durations thus computed, and the longest lower-bound on the range.

Table 1 Mean and range (computed over 22 words for each vowel) of the *shortest*, vowel-nucleus durations across the 10 speakers.

Vowel nucleus	*Mean duration* (*msec*)	*Range of durations* (*msec*)
/i/	137	110 - 205
/e/	183	150 - 270
/a/	129	100 - 170
/o/	192	180 - 215
/ɯ/	126	110 - 180

In Appendix A are listed the 22 words thus selected for each vowel. As shown in those lists, there does not appear any instance of a vowel in the quasi-neutral context "hV" amongst the top 22 words. Indeed, of the 110 words selected, 89 contain a long-vowel "V–" and 21 contain a double-vowel "VV". Moreover, only one word (W0177 in the list for /i/) contains the vocalic nucleus of interest in word-initial position, while 23 words and 86 words contain the vowel nucleus in word-medial and word-final position, respectively. For the vowel /o/ in particular, all 22 words contain the vocalic nucleus in word-final position, and as a lengthened vowel (denoted by "V–"). Although these trends may to a certain extent influence our acoustic-phonetic measurements, it is important to bear in mind our initial criteria for selecting the words, i.e., consistently lengthiest, and therefore presumably most steady-state and least coarticulated, vowel nuclei.

At present, owing to time constraints and the labour-intensive nature of the task, we have completed the formant measurements for five of the 10 speakers. These five speakers, who will hereafter be referred to by their speaker-codes (see below), were of the following age-groups at the time of recording: 20-29 years (S0003 and S0041), 30-39 years (S0001 and S0010), and 40-49 years (S0015). In the next section we describe the methods and results of formant estimation within the vocalic nuclei of the words recorded by these five speakers.

§3 Formant Estimation

Our method of formant estimation proceeded by

using the information provided in each segmentation file, to isolate the speech waveform in only the vocalic nucleus of interest in each recorded word. The speech waveform, sampled at 16kHz, was then subjected to selective linear-prediction (SLP) analysis[18] within the frequency range [0,5]kHz, with the following, default analysis conditions (some of which would later be selectively refined in order to obtain more reliable formants, as described below): coefficient of first-order pre-emphasis PE=0.98, SLP-order M=14, frame-length FL=32msec (512 samples), frame-advance FA=8msec (128 samples), and an FFT-order of 10 in computing the spectrum required in SLP analysis.

A first approximation of the location of the most steady-state part of the vocalic nucleus was then obtained by using a cepstral measure of inter-frame variance[14] (IFV), in each consecutive group of 5 adjacent frames. In particular, the index-weighted or NDPS (Negative Derivative of Phase Spectrum) cepstral distance measure[19] was used to compute the variance in each group of 5 adjacent SLP-cepstra (of order 14), with a frame-group advance of 1 frame; and the group of 5 frames with the lowest variance was taken as the initial estimate of the steady-state within the vowel nucleus. The SLP poles were then found for each of the 5 frames independently by solving for the roots of the polynomial defined by the SLP autoregressive coefficients, and some simple heuristics were applied in order to choose likely formant candidates from amongst the poles. The heuristics comprised simply an allowable frequency range (i.e., a lower and an upper limit) for each of the first four formants, and an upper limit for each formant bandwidth. Using acoustic-phonetic knowledge, these 12 parameters were initially set to default values on a per-vowel basis, and later refined for each speaker according to the subsequent stages of formant correction to be described next.

Indeed, our initial estimates of the steady-state part of each vocalic nucleus, and of the formants in each of those 5 frames, were corrected with the help of formant sequence charts — a spectrogram-like display of the formant candidates, showing the formant frequency patterns in all 5 frames and 22 words of each vowel of each speaker (see Appendix B for the final sequence charts in which the 5 vowels are concatenated for each speaker). The following two criteria were used to visually identify formants in need of correction: (i) inter-frame continuity of the formant frequencies in each steady-state, and (ii) inter-word consistency of the formant frequencies across all 22 words for each vowel of each speaker. Adopting a semi-supervised procedure to correct or to improve the formant measurements in each vowel nucleus, the following parameters were repeatedly adjusted manually:

- the choice of the 5 steady-state frames within the vowel nucleus;
- the SLP order of analysis M;
- the pre-emphasis coefficient PE;
- the frame-advance FA;
- the formant frequency and bandwidth limits used in the heuristic selection of formants from amongst the SLP poles;

and at each step, a new SLP analysis was performed across the entire vowel nucleus. With the help of a computer display showing a quasi-spectrogram of the SLP poles and an updated formant sequence chart, those parameter adjustments continued until our two criteria were satisfied to the best of our abilities.

In particular, of the total of 550 vocalic nuclei analysed (5 speakers × 5 vowels × 22 words), the IFV-computed location of the steady-state was manually shifted in 188 (or 34% of) cases; the order of SLP analysis M which was equal to 14 by default, was modified (within the range 10 to 17) in 98 (or 18% of) cases; the coefficient of pre-emphasis PE which was equal to 0.98 by default, was reduced (to as far as 0.0 in one case) in 95 (or 17% of) cases; and the frame-advance FA which was equal to 8msec (128 samples) by default, was reduced to 5msec (80 samples) in 60 (or 11% of) cases, and to 3.25msec (52 samples) in one particularly difficult case (the vowel /i/ in W1287 of S0015). These parameters needed to be manually adjusted in order to combat the problem of ill-defined formants (either split, missing, or merged formants, or unsteady formant

trajectories); as stated earlier, our criteria were aimed at obtaining the most consistent set of formants across the 5 frames and 22 nuclei for each vowel of each speaker.

In Appendix B are shown the vowel sequence charts finally obtained, for each of the 5 speakers. Those charts reveal that despite all our efforts, there does remain some noticeable variability in certain of the formant sequences — particularly in F3 and F4 of /i/, in F4 of /a/, and in F2 of /ɯ/. Some of that variability, especially in the more difficult to measure, higher formants, may be partly attributed to measurement noise; on the other hand, it is difficult without further, detailed analyses, to separate the influences of frame-to-frame measurement noise, genuine intra-speaker variability, and the potential contribution of inevitable coarticulatory effects from adjacent phonetic segments (or indeed the lack of any consistent phonetic influence in the majority of the vowels which appear in word-final position – suspected to be a particularly strong factor in the variability observed in F2 of /ɯ/[20]). Nevertheless, the sequence charts in Appendix B do portray a reasonably consistent dataset of formant frequencies of the five Japanese vowels.

Supporting that view are the more familiar, F1-F2 and F2-F3 vowel formant planes shown in Appendix C. Indeed, the speaker-pooled vowel clusters are well separated on the F1-F2 plane, where the cluster for the vowel /o/ appears particularly compact (despite the variability expected owing to its word-final position). Furthermore, in support of accumulating evidence[21,22,14,4], there does appear to be a greater amount of inter-speaker variability in F2 of the front vowels, and in F3 of most of the vowels. Indeed, it is interesting to note (cf. **Table 2**) that the per-vowel standard-deviations of the speaker-pooled distribution of formants are substantially higher in the F3 and F4 of all the vowels, and in the F2 of the mid- and high-front vowels.

Our discussions thus far, and indeed our main criteria for securing reasonable measurements, have centred on the formant frequencies, which after all are regarded as the most important acoustic-phonetic parameters for vocalic speech sounds. However, as motivated in our Introduction, we have also retained the *bandwidths* of the SLP poles selected as formants. Admittedly, the formant bandwidths are difficult to measure reliably, and their frame-to-frame variability can also be expected to have been exacerbated by our pitch-asynchronous (or fixed frame-rate) analysis. Nevertheless, our experience of formant measurements suggests that a dataset with sufficient depth (implying number of frames and repetitions of vocalic nuclei) can yield per-vowel *mean* bandwidths which can in turn be used in our method of inversion to yield unique and reasonably realistic VT area-functions[14]; and indeed, our very recent results using the current dataset of formants have confirmed that expectation[13].

In **Fig.1** are plotted the per-vowel mean formant bandwidths obtained by averaging over all 5 frames of the 22 vocalic nuclei of all 5 speakers. As the

Table 2 Standard-deviations (in Hz) of each of the first four formant frequencies' speaker-pooled distributions per vowel.

Vowel	σ_{F1}	σ_{F2}	σ_{F3}	σ_{F4}
/i/	30	133	162	214
/e/	36	111	152	243
/a/	62	75	290	265
/o/	24	49	226	168
/ɯ/	28	141	172	169

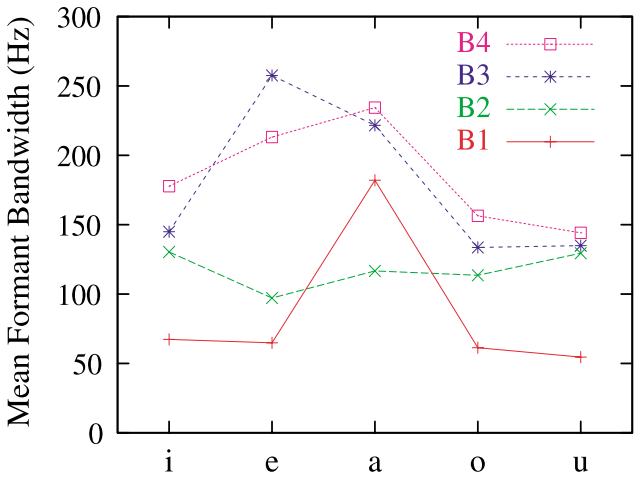


Fig.1 Per-vowel mean formant bandwidths, each computed over 550 tokens (5 speakers×22 words×5 frames).

measurements are of natural speech, these mean bandwidths presumably include the effects of all types of acoustic energy losses in the human vocal-tract, including the lip-radiation, internal losses due to acoustic viscosity, heat-conduction and wall-vibration, and an averaged contribution from the periodic glottal flow. As expected from previous studies of bandwidths measured in natural speech[23,24], the higher formants generally have higher bandwidths. Vowel-specific trends such as the high B1 of the open vowel /a/, the high B3 of the mid-front vowel /e/, and the low B1, B3 and B4 of /o/ and /ɯ/, would need to be verified against similar data for the vowels of Japanese and other languages.

§4 Electronic copy of Formant Data

The semi-supervised method of formant measurement described in this paper, involves a combination of automated, computer analysis of speech, and manual application of phonetic knowledge on the part of the speech scientist. It is in this sense that the truly reliable estimation of formants, even within the presumably more easily measured, steady-states of vocalic nuclei, remains to this day as much an art as a science. Our painstakingly-constructed corpus of the first four formants of vowel steady-states recorded by 5 adult, male speakers of Japanese, is therefore offered to the research community, at the following address:

http://www.etl.go.jp/etl/onsei/formant.html

Acknowledgement

The first author gratefully thanks Dr Frantz Clermont for his encouragements during the Japanese spring of 1999 when the formant measurements were initiated.

References

1) G.E.Peterson & H.L.Barney, “Control Methods Used in a Study of the Vowels,” *J. Acoust. Soc. Am.* **24**, 175-184 (1952).

2) R.L.Watrous, “Current status of Peterson-Barney vowel formant data,” *J. Acoust. Soc. Am.* **89**, 2459-2460 (1991).

3) F.Clermont, “Multi-speaker formant data on the Australian English vowels: A tribute to J.R.L.Bernard’s (1967) pioneering research”, *Proc. 6th Australian Int. Conf. on Speech Science & Tech.*, 145-150 (1996).

4) M.Barlow & F.Clermont, “Download versions of the Bernard data,” ADFA Speech Process. Lab.: John Bernard Database, http://www.cs.adfa.edu.au/speech/JB

5) J.R.L.Bernard, “Some measurements of some sounds of Australian English”, Doctoral Thesis, The University of Sydney (1967).

6) G.Fant, “Formant Frequencies of Swedish Vowels”, in *Speech Sounds and Features* (MIT Press), 94-99 (1973).

7) H.Fujisaki & K.Yoshimune, “Analysis, normalization and recognition of sustained Japanese vowels,” *Annual Report of the Engineering Res. Inst.* **29**, Faculty of Eng., Univ. of Tokyo (1970).

8) L.C.W.Pols, H.R.C.Tromp, & R.Plomp, “Frequency analysis of Dutch vowels from 50 male speakers,” *J. Acoust. Soc. Am.* **53**, 1093-1101 (1973).

9) D.J.P.J.van Nierop, L.C.W.Pols, & R.Plomp, “Frequency Analysis of Dutch Vowels from 25 Female Speakers,” *Acustica* **29**, 110-118 (1973).

10) D.J.Broad & H.Wakita, “Piecewise-planar representation of vowel formant frequencies,” *J. Acoust. Soc. Am.* **62**, 1467-1473 (1977).

11) D.J.Broad, “Piecewise-planar vowel formant distributions across speakers,” *J. Acoust. Soc. Am.* **69**, 1423-1429 (1981).

12) J.Hillenbrand, L.A.Getty, M.J.Clark, & K.Wheeler, “Acoustic characteristics of American English vowels,” *J. Acoust. Soc. Am.* **97**, 3099-3111 (1995).

13) P.Mokhtari, F.Clermont, & K.Tanaka, “Toward an acoustic-articulatory model of inter-speaker variability”, to appear in *Proc. Int. Conf. on Spoken Lang. Process*, Beijing, China (2000).

14) P.Mokhtari, “An acoustic-phonetic and articulatory study of speech-speaker dichotomy”, PhD Thesis, The University of New South Wales, Canberra, Australia (1998).

15) P.Mermelstein & M.R.Schroeder, “Determination of smoothed cross-sectional area functions of the vocal tract from formant frequencies”, *Proc. 5th Int. Congress on Acoustics*, Liège, Paper A24 (1965).

16) P.Mokhtari & F.Clermont, “New perspectives on linear-

prediction modelling of the vocal-tract : uniqueness, formant-dependence and shape parameterisation", to appear in *Proc. Australian Int. Conf. on Speech Science and Tec*., Canberra, Australia (2000).

17) S.Hayamizu, K.Tanaka, S.Yokoyama, & K.Ohta, "Generation of VCV/CVC Balanced Word Sets for Speech Data Base", *Bul. Electrotech. Lab.* **49** (10), 803-834 (1985).

18) J.D.Markel & A.H.Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, Heidelberg, New-York (1976).

19) B.Yegnanarayana & D.R.Reddy, "A distance measure based on the derivative of linear prediction phase spectrum", *Proc. Int. Conf. on Acoust., Speech, and Sig. Process.*, 744-747 (1979).

20) H.Fujisaki, Y.Sato, & T.Yamakura, "Recognition of semivowels of a number of speakers based on a model of coarticulation", *Proc. Spring Meeting of Acoust. Soc. of Japan*, 269-270 (1975).

21) P.Mokhtari & F.Clermont, "Contributions of selected spectral regions to vowel classification accuracy", *Proc. 3rd Int. Conf. on Spoken Lang. Process.*, 1923-1926 (1994).

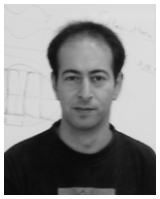
22) P.Mokhtari & F.Clermont, "A methodology for investigating vowel-speaker interactions in the acoustic-phonetic domain", *Proc. 6th Australian Int. Conf. on Speech Science and Tech.*, 127-132 (1996).

23) B.P.Bogert, "On the Band Width of Vowel Formants", *J. Acoust. Soc. Am.* **25**, 791-792 (1953).

24) H.K.Dunn, "Methods of Measuring Vowel Formant Bandwidths", *J. Acoust. Soc. Am.* **33**, 1737-1746 (1961).

(Accepted September 8)

著 者 紹 介



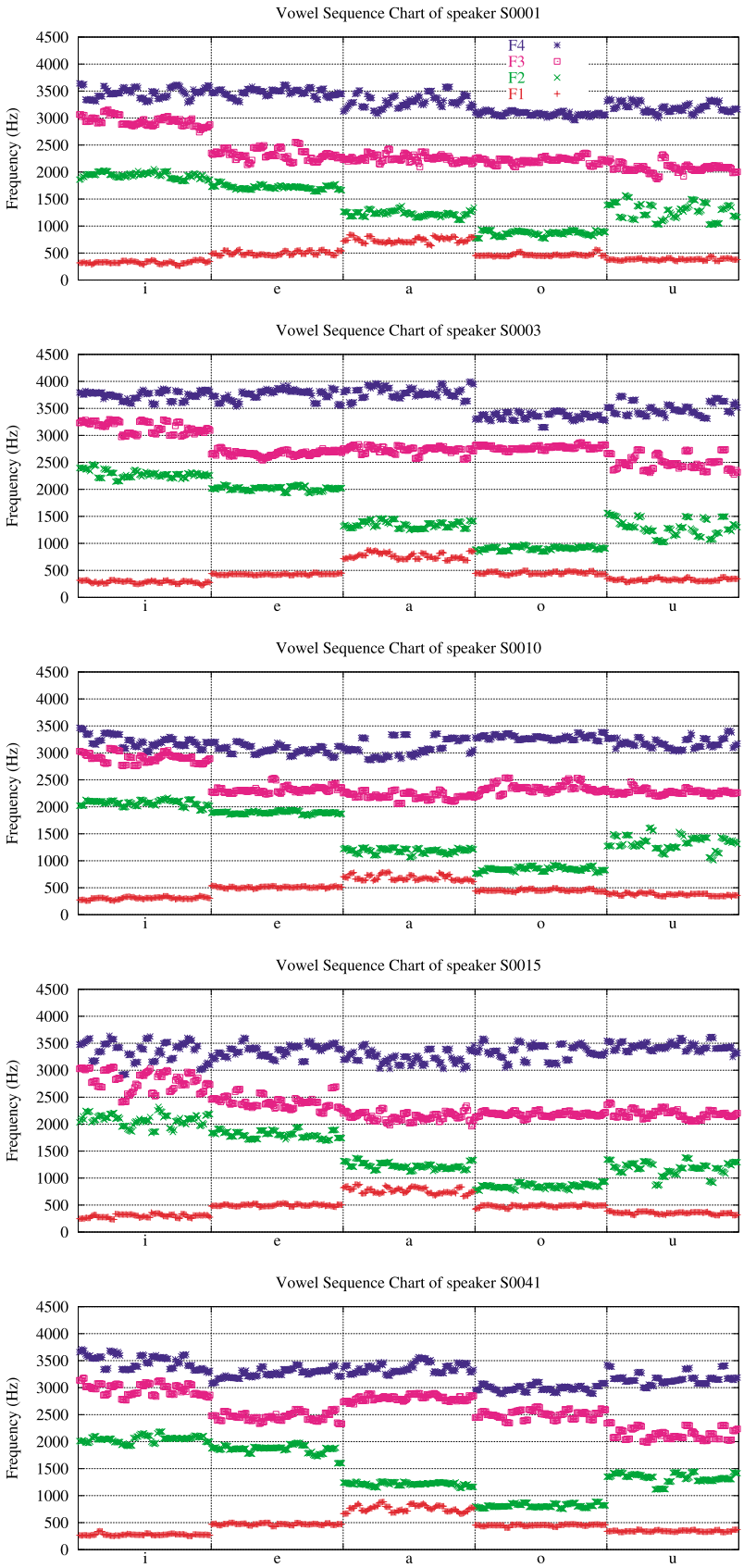
パーハム・モクタリ
Parham MOKHTARI
知能情報部　音声信号処理ラボ
E-mail:parham@etl.go.jp
音声信号処理・音声合成の研究に従事。最近のテーマは，話者特徴のモデル化，音質変換。

田中　和世
Kazuyo TANAKA
知能情報部　音声信号処理ラボ
E-mail:ktanaka@etl.go.jp
音声情報処理の研究に従事。最近のテーマは，言語系に依存しない音声記号系を用いた音声処理。

Appendix A - List of Selected Words

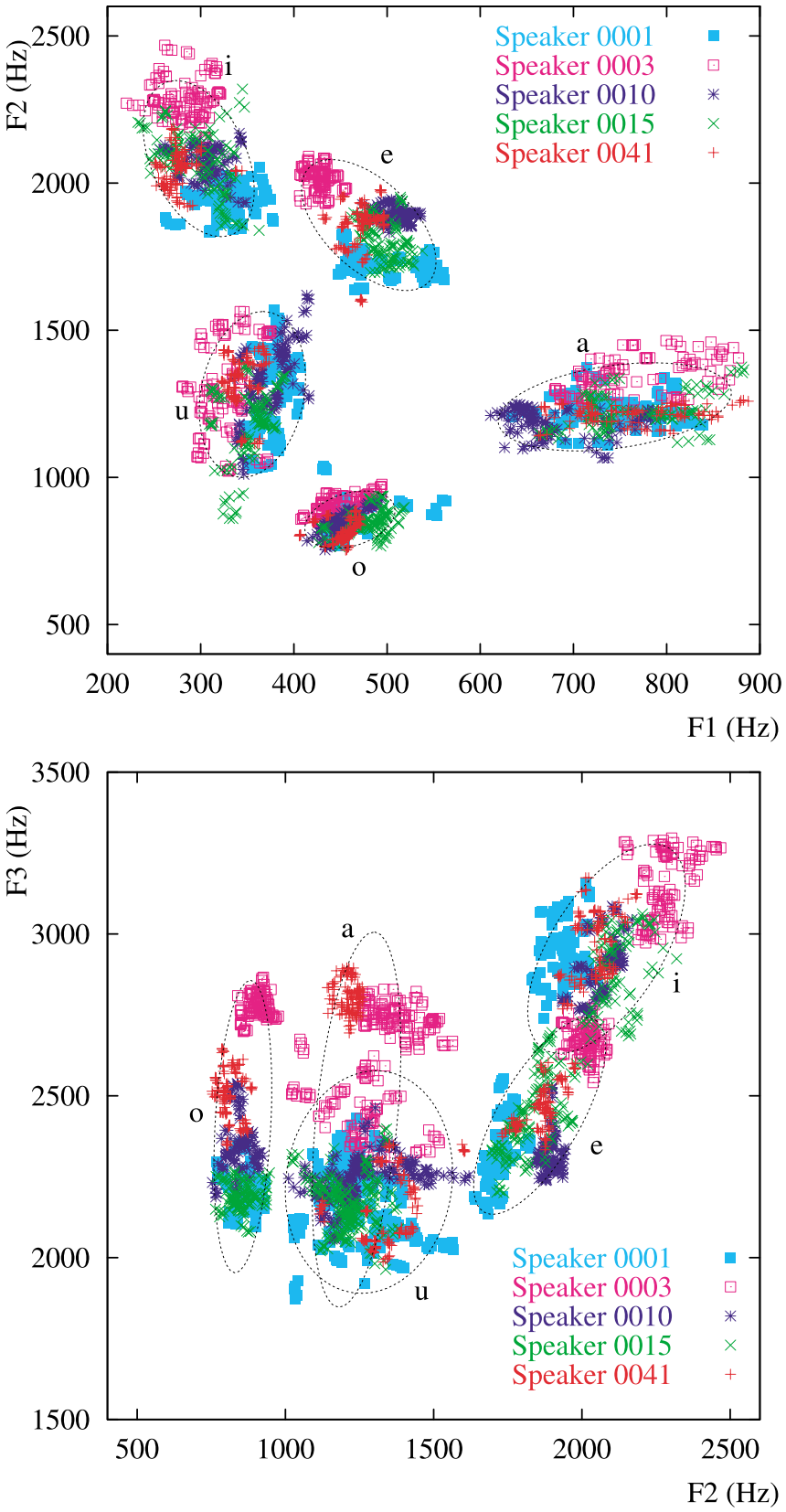
List of all 110 words (their codes and romanised phonetic representations) selected from the “ETL-WDI and II” dataset. See Section 2 for a description of the word-selection methods and criteria, which aimed to identify the lengthiest and potentially most steady-state vowel nuclei (shown below in bold font).

	/ i /		/ e /		/ a /		/ o /		/ɯ/	
	Code	*Word*	*Code*	*Word*	*Code*	*Word*	*Code*	*Word*	*Code*	*Word*
1	W0483	za**ii**	W1521	yo**e**-	W1113	edit**a**-	W1218	jah**o**-	W1092	chiNts**u**-
2	W0240	ku**ii**ji	W1143	gu**e**-	W1063	b**aa**i	W1126	gaz**o**-	W1526	yosh**u**-
3	W0519	bod**i**-	W1514	ya**e**-	W1249	k**aa**tsu	W1449	shud**o**-	W1006	us**uu**su
4	W0795	mob**i**-ruyu	W1152	gyo**e**-	W0393	shir**aa**e	W1238	jud**o**-	W0898	ryosh**u**-
5	W0946	shuj**ii**	W1522	yoh**e**-	W1404	rez**a**-	W1146	guk**o**-	W1010	uz**uu**zu
6	W0177	**ii**ki	W1157	gyos**e**-	W0924	shaw**a**-	W0677	jad**o**-	W1300	kyos**u**-
7	W0212	kar**ii**o	W1244	jur**e**-	W0871	paud**a**-	W1151	gyod**o**-	W1447	shoy**u**-
8	W1287	kuchiosh**i**-	W1148	gut**e**-	W1119	fakut**a**-	W1108	doj**o**-	W1241	juky**u**-
9	W1479	suzush**i**-	W1450	sh**ue**-	W0135	had**aa**i	W0901	ryoy**o**-	W0398	shuk**uu**
10	W0612	haebaesh**i**-	W1452	shuh**e**-	W1403	reshi-b**a**-	W1534	yuj**o**-	W0927	shib**uu**chiwa
11	W1118	eruesud**i**-	W0900	ryot**e**-	W0334	poiNt**a**-	W1301	kyoz**o**-	W1528	yoy**u**-
12	W1027	yu-tirit**i**-	W1089	chat**e**-	W0872	paw**a**-	W1448	shub**o**-	W1537	zay**u**-
13	W1267	kewash**i**-	W0531	buz**e**-	W0610	h**aa**ri	W1127	ged**o**-	W1158	gyosh**u**-
14	W1067	beb**i**-	W0998	uNt**e**-	W0758	kyaNp**a**-	W0771	kyoy**o**-	W0111	geqts**u**-
15	W1371	ob**ii**wai	W1421	sha**e**-	W0498	ak**aa**za	W1424	shas**o**-	W0678	jash**u**-
16	W1116	enuj**i**-	W1437	sho**e**-	W0439	top**a**-zu	W1296	kyod**o**-	W0693	joy**u**-
17	W0683	ji-enup**i**-	W0616	haijitsus**e**-	W1294	kyasut**a**-	W1195	huch**o**-	W1340	muy**u**-byo-
18	W1355	nitsukawash**i**-	W1444	shos**e**-	W1106	dep**a**-to	W1420	shad**o**-	W0885	ramu**u**-ru
19	W1532	yu**ii**tsu	W0751	kuNs**e**-	W0385	shep**a**-do	W1159	gyot**o**-	W1341	myak**uu**tsu
20	W0831	niginigish**i**-	W1137	git**e**-	W1510	y**aa**wase	W1245	jury**o**-	W0083	doNts**u**-
21	W0981	terepash**i**-	W0715	kar**ee**da	W0099	faNf**a**-re	W1451	shugy**o**-	W1185	hoky**u**-
22	W0966	tadotadosh**i**-	W1454	shus**e**-	W1442	shorud**a**-	W1075	biz**o**-	W0582	giday**u**-



Appendix B - Vowel Sequence Charts

Vowel sequence charts for each of the 5 speakers, showing the distribution of the first four formant frequencies {F1, F2, F3, F4}. Within each vowel category, the horizontal sequence of 110 formant patterns comprises 5 adjacent steady-state frames × 22 vocalic nuclei.



Appendix C - Vowel Formant Planes
The 5 speakers' vowel formant planes F1-F2 and F2-F3, with a 2-σ ellipse drawn around each vowel cluster.