EFFECTIVE F2 AS A PARAMETER IN JAPANESE FORENSIC SPEAKER IDENTIFICATION

Yuko Kinoshita
Phonetics Laboratory, Linguistic Program,
The Australian National University

ABSTRACT The possibility of using effective F2 as a parameter in forensic speaker identification is demonstrated. 11 male Japanese speakers were recorded, and the formants of their 5 Japanese accented short vowels were measured. The potential of effective F2 as a possible forensic speaker identification parameter was evaluated by F-ratio. The results showed that effective F2 of /e/ produced considerably a higher F-ratio than individual F2 and F3 of the same vowel, although transforming into effective F2 did not improve the F-ratio of the other vowels.

INTRODUCTION

This paper studies the possibility of the use of effective F2 in Japanese as a parameter for forensic speaker identification. It is well known that formant structures contain considerable information on a speaker's identity. In fact, formants are one of the most widely used parameters for testing forensic speaker identification (see, e.g. Greisbach et al. 1995, Jessen 1997, Nolan 1983, Rose 1999a; 1999b). The formant pattern is the acoustic reflex of the supralaryngeal vocal tract configuration. The size and shape of the supralaryngeal area are determined by two factors: which segment is being produced, and the speaker's anatomical features. Differences in the formant patterns of the same sound are thus considered to reflect speakers' anatomical differences, making the formant pattern one possible parameter in speaker identification.

Formant centre frequencies are one of the most typically measured parameters in real forensic cases (Rose 1999b: 4). Although there are other parameters which also have a potential to distinguish speakers, formant pattern has been preferred in forensic situations over other parameters, such as the cepstrum, for two reasons. Firstly, formant patterns can be interpreted in relation to auditory features. The first step in any forensic speaker identification procedure is always an auditory analysis. Auditory analyses are indispensable in both the evaluation of the comparability of samples and the selection of the potential parameters. An analysis of formant patterns enables us to quantify and to discuss the results of the auditory analysis acoustically, whereas it is much more difficult to relate the automatic recognition parameters to auditory impression (but see Clermont and Itahashi (2000)). Additionally, the parameters for automatic recognition are based on mathematically more complex notions than formant patterns, making it harder for juries to understand. In forensic situations, the results of an analysis must be comprehensible by juries (who are unlikely to have a phonetics background!).

In forensic speaker identification, formants are usually analysed individually, producing a separate evaluation for each formant. This method is, however, less than ideal. Because of the complexity of human speech production, an evaluation based on a single formant can never be sufficient to make a statement on a given speaker's identity, so that multiple parameters must be taken into consideration. Comparing the similarity of multiple values is, however, far more complex than comparing the similarity of single values, and analyses are thus not an easy task. The more appropriate approach seems, therefore, to measure the formants and then statistically reduce their dimensionality. In this way, the simultaneous comparison of many parameters, which is absolutely necessary to profile a speaker, can be performed considerably more easily. Effective F2, the focus of this study, is a value which approximates the auditory differences of speech sounds. What makes effective F2 an attractive parameter is the fact that it is formulated by incorporating information from the first three formants into a single figure. In other words, it serves as a dimensionality-reducing function. Additionally, effective F2 may be an easier concept for juries to understand than the abstract figures which are the result of complicated statistical procedures. As has been noted in the discussion of the preference towards formant patterns, understandability for juries

should not be neglected in forensic speaker identification. The purpose of the expert witness is to present the evaluation of evidence as a part of material for use by juries in making of judgements.

The reasons discussed above suggest that effective F2 might have potential in forensic speaker identification. Since no investigation has been made on the use of this parameter, this study explores the possibility of effective F2 as a parameter for speaker identification, by comparing it with the results of analysis of conventional vowel formants.

PROCEDURE

Recording One of the reasons that forensic speaker identification is much more difficult than automatic one is the lack of control over data. Having natural speech samples as data is, therefore, one of the desiderata in forensic speaker identification experiments. The recordings in this study were done using tasks carefully designed for the elicitation of natural speech. In these tasks, the informants were provided with a map and an information sheet on 4 people. The map contains 3 bus routes and names of shops and buildings. The information sheet consists of 4 people's jobs, personalities, and favourite foods. The informants were requested to answer the questions such as "Where does the route A bus stop?" or "What kind of job does person A do?," referring to the given materials. The map and the information sheet were designed to contain examples of all 5 Japanese short vowel phonemes occurring on the pitch accented syllable, 5 times each. The linguistic contents of the corpus are summarised in table 1.

/a/	*hanaya* 'florist', *panya* 'bakery', *sakata* '(name)', *sobaya* 'noodle shop', *panyano* 'of bakery'
/i/	*jinja* 'shrine', *jibika* 'otolaryngology', *kobijutsu* 'antique', *sushiya* 'sushi bar', *sanwaginkoo* 'Sanwa bank'
/u/	*nikuya* 'butcher', *tokushima* '(name)', *kaguten* 'furniture shop', *doobutsuen* 'zoo', *kurita* '(name)'
/e/	*Nemoto* '(name)', *terebi* 'TV', *kitadeguchi* 'north exit', *kitadeguchi* 'north exit', *minamideguchi* 'south exit'
/o/	*Kinoshita* '(name)', *toshokan* 'library', *hoteru* 'hotel', *honya* 'book shop', *toposu* '(name of shop)'

Table 1. Words included in the corpus of natural speech. The accented segments are underlined.

The informants for this study were 11 male native speakers of Japanese. The informants were requested to repeat these tasks once, and two recording sessions, separated by two weeks, were held for each speaker. Exactly the same process was followed at the both recording sessions. The recording was carried out in the studio of the Phonetics laboratory at ANU.

Measurements The recordings were digitised at 16 kHz and analysed with the CSL sound analysis software package. Formants of the short accented vowels were sampled in the middle of the vowel duration. This measuring point was chosen to minimise the effect of the adjacent segments on the measurements of the target vowels. Each of the vowel / formant combinations consists of 20 samples (5 words * 2 repeats * 2 recording sessions).

EFFECTIVE F2

Effective F2 was originally used as a perceptually based transform for optimal separation of vowels in languages with high- and mid- front rounded vowels, such as Swedish (Fant 1973). Experimenting on Swedish vowels, Fant reports that back vowels can be approximated well using only natural F1 and F2, whereas for some other vowels, front-rounded vowels in particular, F3 and formants above are also relevant. Fant thus decided to take F3 into account for the separation of vowels. As result, a value located between natural F2 and F3 was formulated and proposed as effective F2. The vowel mapping based on F1 and effective F2 more successfully separated the vowels, which overlap heavily in the natural F1/F2 plane. Fant's formula is shown at (1)

(1) Effective F2 = F2+1/2(F3-F2)*(F2-F1)/(F3-F1) (Fant 1973:52)

Table 2 presents the results of the effective F2 calculation using this formula.

Speaker	/a/		/i/		/u/		/e/		/o/	
	mean	sd.	mean	sd.	mean	sd.	mean	sd.	mean	sd.
AA	1519.8	124.13	2352.8	115.62	1703.9	122.06	2037.7	80.92	1246.9	256.06
HA	1613.1	189.65	2705.6	145.47	1885.4	233.06	2458	91.027	1271.9	143.62
JN	1674.4	109.56	2223.2	193.88	1820.7	169.38	1996.6	81.295	1350.4	178.37
KA	1782.4	161.55	2393.9	194.49	1906.8	103.67	2230.2	142.57	1339.2	232.75
KF	1795.4	199.38	2541.2	155.48	1893.6	237.18	2337.4	98.615	1425.2	164.94
KH	1674.6	170.12	2371.4	133.68	1845.9	179.19	2167.1	56.72	1421.9	327.79
KO	1643.5	169.9	2242.8	191.94	1944.1	188.84	2166.7	69.081	1346.6	206.07
MN	1656.4	181.09	2663.2	150.97	1738.1	214.47	2391.3	125.65	1384	139.95
TN	1799.9	159.07	2431.9	135.94	1803.8	306.13	2291.4	71.765	1525	267.74
TS	1508.7	223.75	2462.9	92.819	1846.1	152.37	2142.8	70.909	1406.2	227.4
TY	1479.4	132.54	2186.2	128.02	1582.7	144.17	2061.2	72.717	1161.7	228.88
Mean	*1649.8*	*165.52*	*2415.9*	*148.94*	*1815.5*	*186.41*	*2207.3*	*87.39*	*1352.6*	*215.78*
Sd.of M		*113.9*		*170.2*		*105.1*		*149.4*		*99.0*

Table 2 Mean and standard deviation of effective F2 for the short vowels of 11 Japanese males. Left-hand column shows speakers. Vowels are shown in the top row. (n=20, as long as there is no missing value.)

allophone of this vowel is usually transcribed as [ɯ].

Now, it is of interest to see how effective F2 works in Japanese. Figure 1 shows the scattergrams of each speaker's mean F1, F2 and effective F2 for five Japanese vowels. Natural F1 and F2 are plotted in the left figure and natural F1 and effective F2 are plotted in the right. It should be noted here that the phoneme /u/ in Japanese is not rounded. The main

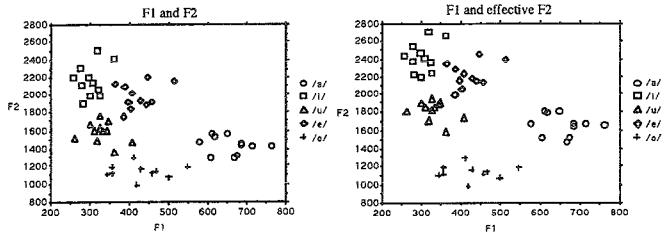

Figure 1. Plots of F1 vs. F2 (left) and effective F2 (right) for 11 speakers' five Japanese short accented vowels. Values are in Hz. Each symbol shows the mean of the 5 vowels for a given speaker.

Figure 1 shows that, although the vowels are quite well separated even with natural F2, effective F2 does seem to improve the separation between back vowels /o/ and /u/, and /o/ and /a/. Improvement in the separation of back vowels appears not to agree with Fant's observation (he reported effective F2 separated front-rounded vowels particularly well). Effective F2 maybe separates rounded vowels in general well, but not just front-rounded vowels. In Figure 1, it seems that only the rounded vowel /o/ was kept low whereas the other vowels /u/ and /a/ raised by the calculation of effective F2.

The fact that Japanese has a better separation of vowels than Swedish is not surprising considering that Japanese has only 5 vowel phonemes, whereas Swedish has 14. Even if the vowels of these two languages have a similarly sized spread, Japanese is less likely to have overlaps. It must be noted, however, that having a smaller set of vowel inventories does not necessarily mean that each vowel of the language will maximally utilise the given vowel space.

ANOVA

Between- and Within-speaker variance Speaker identification relies on the notion that between-speaker variation is larger than the within-speaker variation. Acoustically no one says anything in exactly the same way twice, so there is always within-speaker variation, even in the most similar utterances made by a

single speaker in a short period of time. This variation is, however, smaller than the variation between two different speakers in most cases. If within-speaker variation of one speaker is as large as the variation between the speaker and a second speaker, then it would not be possible to distinguish these two speakers. Conversely, when any speaker identification parameter shows small within-speaker variation and large between-speaker variation, the analysis based on the parameter is liable to be useful in speaker identification. Small within-speaker variation and large between-speaker variation are therefore regarded as essential criteria for forensic phonetic acoustic parameters (Nolan, 1983:11).

As this study aims to investigate the potential utility of effective F2 in forensic investigations, the statistical method employed in this study needs to be the one which can reveal the size of both within- and between-speaker variation. The results of measurements are hence analysed with ANOVA. ANOVA is a statistical method for comparing a number of groups based on the mean values within those groups, and evaluates whether are any significant differences among those groups. In this study, 11 groups were compared, as 11 speakers were employed as informants.

The results of ANOVA are presented in terms of F-ratios. The F-ratio is the ratio of the between group variation to the within group variation (Hatch and Lazaraton, 1991:315). In the present study, as each speaker comprises one group, within- and between-group variation can be paraphrased as within- and between-speaker variation. So, for instance, an F-ratio of 3 means that between-speaker variation is three times larger than within-speaker variation. An F-ratio under 1 means within-speaker variation is larger than between-speaker variation for the parameter, making it of no use for speaker identification. The larger the F-ratio is, *ceteris paribus*, the more powerful the associated parameter is. Thus our interest in this study can be expressed in terms of determining how much larger (or smaller) the F-ratio of effective F2 is compared to other formants.

RESULTS

Firstly, the F-ratios of both natural formants and effective F2 obtained from one way factorial ANOVA are presented presented in Table 3 below. F-ratios over 20 are marked by shading.

	F1	F2	F3	F4	F'2
/a/	8	9	9	5	8.8
/i/	7	26	10	8	21.1
/u/	8	3	10	3	2.9
/e/	11	31	27	8	43.3
/o/	12	3	6	7	3.6

Table 3. F-ratios for all vowel / formant combinations.

Table 3 shows that, for natural formants, F2 of /i/, and F2 and F3 of /e/ have a higher F-ratio than the other formants (26, 31, and 27 respectively). As for effective F2, /e/ is shown to have a considerably larger F ratio than any other formants. The calculation of the effective F2 did not improve the F-ratio of the other vowels, however. The other three vowels, namely /a/, /u/, and /o/, did not show a significant difference between the F-ratios of natural F2 and effective F2. As for the /i/ vowel, although the F-ratio of effective F2 (21.1) seems large, it is notably smaller than that of natural F2. The generalisation "effective F2 is the better measure for distinguishing speakers irrespective of vowel" is, therefore, not valid. The value of 43.4 which the effective F2 of /e/ produced was the highest F-ratio by far amongst all of the results collected so far, however, and also the /e/ vowel was found to be one of the promising vowels as a parameter from the study of the natural formants. This result seems thus more than just anomalous. The table of F-ratios for natural formants above (table 2) shows that /e/ is the only vowel which has relatively high F-ratios for F1, F2 and F3. Incorporating originally powerful parameters may have produced an even more powerful parameter. For the other vowels which did not have high F-ratios for raw F2 and F3, on the contrary, the effective F2 calculation did not improve the F-ratios. This may have been especially true for /i/. Combining the strong parameter F2 (F-ratio 26) with less powerful parameters F3 (F-ratio 10) may have resulted in a lower F-ratio (21.1).

Comparison with previous findings Analysis using ANOVA revealed that the effective F2 of /e/ yielded an F-ratio of 43.3. How 'large' this value is can be appreciated by comparison with previous findings. This section compares the F-ratio to that of the previous findings to have a more objective view. Table 4 summarises previous studies.

	CONDITION AND RESULTS
Wolf (1971)	-read out 6 sentences (21 American English speakers) -F2: 'cash' 46.6, 'the' 44.6, 'papa' 19.0
Nolan (1983)	-word list reading (15 British English speaking 17 years old male speakers) -/l/ F1 17.7, F2 21.6, F3 77.8 -/r/ F1 46.4, F2 59.4, F3 216.9
Rose (1999)	-'hello' in 6 different ways (6 similar-sounding Australian English speakers) -/l/ F2 43.5, offglide of /ou/ F2 44.1

Table 4. Summary of F-ratios in preceding research

The F-ratio for natural formants of this study is considerably lower than that of the studies summarised above. Two explanations for these low F-ratios are proposed. Firstly, recording styles differ significantly. The data used in the current study were elicited from natural speech, whereas Nolan (1983) and Wolf (1971) used words or sentences from readings for their data collection. The effect of the difference in speech styles is significant, as natural speech has considerably larger within-speaker variation than readout speech (Kinoshita 1998). Rose (1999a) used the word 'hello,' asking the speakers to utter the word in 6 different ways, such as answering the phone or announcing their arrival home. Although this method would enable making the recording closer to natural speech than readout speech, the utterances are still not fully spontaneous. His data therefore should also be distinguished from the natural speech used in this study. Furthermore, the fact that the experiment in his study was on the vowels embedded in the different contexts, whereas in Rose, all segments are elicited from the same word, "hello," and this probably has contributed to the size difference in within-speaker variation. The lack of control over speech style in this data probably explains its small F-ratio. From the point of view of forensic acoustic investigation, the F-ratios collected in this research, which is based on natural speech, seem to be more realistic values.

Secondly, the linguistic difference -- the Japanese phoneme structure -- should be considered. Japanese has less vowels than English as shown in Figure 2. This difference in the size of potential space for each vowel to occupy may have contributed to the larger within-speaker variation of Japanese speakers. As speakers cannot go beyond the physical limitation of their vocal range, the between-speaker variation may not be affected as much as within-speaker variation. If the influence on within-speaker variation is larger than that on between-speaker variation, the corresponding F-ratios will consequently become smaller.

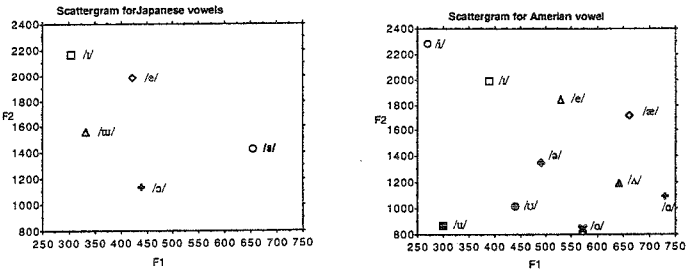

Figure 2. Mean F1 and F2 plots for Japanese and English. The English vowels are based on the 1952 P&B data quoted in Backen, 1996:358),

This observation implies the possibile existence of language specific differences in the realisation of within- and between-speaker variation. If there are any language-specific tendencies regarding within- and between-speaker variation, the validity of the speaker identification involving two (or more) different languages as data is in serious question, as would also be an attempt to apply statistics from one language to another. Nevertheless, the scattergrams of each speaker's mean F1 and F2 (Figure 1) have shown that Japanese vowels do not necessary spread more widely just because they can. To make any conclusive remark on the relationship between the vowel space and F-ratio, comparison of the standard

deviations is also necessary. Comparisons between languages with both similar and different vowel phoneme structure, such as Japanese / Spanish, and Japanese / English, will be an interesting future task

CONCLUSION

The experiment described in this paper has demonstrated the potential of effective F2 as a parameter for speaker identification. The results show that effective F2 can have much higher F-ratios than natural formants, although this is only the case for /e/. In this study, only the /e/ vowel improved its F-ratio by calculating effective F2. The three other vowels, /a/, /u/, and /o/, did not change their value, and the /i/ vowel lowered its F-ratio. The improvement of the F-ratio for /e/ appears significantly large (31 to 43.4) and, with the exception of some of Nolan's results, the F-ratio of effective F2 of /e/, 43.4, is as large as (or even larger than) that of the preceding studies. Considering that these preceding studies had conditions more conducive to obtaining larger F-ratios, effective F2 for /e/ is assumed not just to be more powerful than the other formants, but also its discriminatory power may be significantly stronger. This of course now needs to be tested.

Further, through comparison with previous research, the possibility of language-specific characteristics in the realisation of between-speaker variation was found. This compels us to take extra precaution in the comparison of speech samples which are in different languages.

ACKNOWLEDGEMENT I am very grateful for the reviewers who provided me with many valuable comments.

REFERENCES

Backen, R.J. (1996) Clinical Measurement of Speech and Voice, (Singular Publishing Group Inc, San Diego).

Clermont, F. and Itahashi, S. (2000) "Static and Dynamic vowels in a Cepstro-phonetic sub-space", J. Acoust. Soc.Jpn. 21/4 221-223.

Fant, G. (1973) Speech Sound and Features, (MIT Press: Cambridge) .

Greisbach, R. Esser, O. and Weinstock, C. (1995) "Speaker identification by formant contours", Studies in Forensic Phonetics *BEIPHOL 64* 49-55.

Hatch, E. and Lazaraton, A. (1991) The Research Manual - Design and Statistics for Applied Linguistics, (Heinle & Heinle: Boston).

Jessen, M. (1997) "Speaker-specific information in voice quality parameters", Forensic Linguistics *4*:84-103.

Kinoshita, Y. (1998) "Japanese forensic phonetics: Non-contemporaneous within-speaker variation in natural and read-out speech", In eds 5th International Conference on Spoken Language Processing

Nolan, F. (1983) The Phonetic Bases of Speaker Recognition, (Cambridge University Press: Cambridge).

Rose, P.J. (1999a) "Differences and distinguishability in the acoustic characteristics of *Hello* in voices of similar-sounding speakers", Australian Review of Applied Linguistics *21* 1-42.

Rose, P.J. (1999b) "Long- and short-term within-speaker differences in the formants of hello", Journal of the International Phonetic Association *29* 1-30.

Wolf, J.J. (1971) "Efficient acoustic parameters for speaker recognition", JASA *51* 2044-2055.