TR－H－005

From EMG to Sound Patterns of Vowels : Software

Shinji Maeda

1992. 7. 20

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2　☎07749-5-1011

ATR Human Information Processing Research Laboratories

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

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

From EMG to sound patterns of vowels : Software

Shinji Maeda

Ecole Nationale Supérieure des Télécommunications (ENST)
Département SIGNAL
and
Centre National de la Recherche Scientifique (CNRS)
URA 820

46, rue Barrault, 75634 Paris cedex 13, France

Tel (33 1) 45 81 71 91; Fax (33 1) 45 88 79 35;
Email maeda@sig.enst.fr

abstract

With a few exceptions (Kakita, Fijimura & Honda,1985), EMG data are interpreted with reference to the intended output, such as the phonetic description of utterances spoken by speakers. For a more rigorous interpretation, the data should be also analyzed in terms of the displacement of the articulators and the acoustic patterns. In this paper, we describe our attempts to calculate the formant patterns from EMG responses with an intermediate articulatory model (Maeda, 1990). Since the results are rather scant at this early stage, the main purpose of this report is to document how to use the developed software so it can be used by other researchers. However, I shall describe some of results in order to explain how the software can be used effectively and to demonstrate how useful to evaluate the contribution of tongue and lip muscle activity in terms of articulatory positioning and acoustic pattering, thus filling the gaps between the muscular activity and the acoustics.

1. Introduction

In this report, I describe extension of an articulatory model based on a factor analysis of X-ray film data, developed at CNET in Lannion, France (Maeda, 1979, 1990). The current version has been implemented on a PC at the research unit of CNRS (URA-820) associated with the SIGNAL department of the Ecole Nationale Supérieure des Télécommunications (ENST) in Paris. Specifying the values of the seven positional parameters, such as jaw position, tongue-body position, *etc.*, the model calculates the vocal tract profile and the frontal shape of the lip opening. Acoustic characteristics, such as the transfer function and formant frequencies, and the signal of the corresponding stationary vowel are computed from the calculated tract configurations. Motivated by fairly complete EMG data of the extrinsic tongue muscles and of the lip muscle (orbicularis oris) at ATR, and inspired by the recent work of Honda and Kusakawa on the EMG description of vowels (Honda, 1992; Honda, Kusakawa and Kakita, 1992; Kusakawa, Honda, and Kakita, forthcoming), I have written an interface program in which the averaged EMG responses representing muscular forces are converted into values of the positional model parameters. The complete system, therefore, allows us to study EMG activity in relation to model tract configurations and the resulting acoustic characteristics of vowels.

The attempt to calculate formant frequencies from EMG activities is not new. Already a few researchers (for example, Kakita and Fujimura, 1984) have done such calculations using a tongue model based on a finite element method. My model, however, is purely descriptive. The tract shapes are determined as the sum of proportional influences of the individual articulatory parameters. The model is the result of an arbitrary orthogonal factor analysis (Overall, 1962) on X-ray film data. What I shall propose for the force-to-position conversion is also the weighted sum of the EMG responses of the paired antagonistic muscles. Although we might not expect such a linear model from EMG to sounds to work very well, the preliminary results are rather encouraging. It was possible to obtain reasonable F1-F2 patterns for 11 American English vowels with straightforward corrections of the lip aperture and tongue body position parameters. The lip correction is reasonable since only the EMG response of the closing orbicularis oris muscles is available. For most of back vowels from /u/ to /o/, the EMG specified tongue body position is not sufficiently backed. Consequently, the F2 frequency becomes too high for these back vowels.

An advantage of the linear modeling is its simplicity. Even on a PC, it is possible to calculate the formant variations of utterances fairly quick. In fact, most of the CPU time is consumed by the acoustic calculation (Maeda, 1982) rather than by the calculation of the tract shape and the subsequent area function. The main purpose of this report is to document the developed software so it can be used by other researches, rather than to describe the results which are rather scant at this early stage. However, I shall describe some of results, in order to demonstrate how the software can be used effectively in experiments.

2. General views

All programs are written in C using Borland Turbo C, v2.0, (TC) running on an IBM-PC compatible. The system includes function libraries, their source and include files, and main programs. User ready programs **lam**, **vtcalcs**, and **emglam** (see Appendix B for how to use them) are also included. There are three source files specifically for the EMG-to-sound problem, **combin.c**, **emglam.c**, and **emgfrmnt.c** which are explained in detail. The whole software package is under \maeda\ and its sub-directories. The software package fits on two floppy disks and can be transferred to any IBM-PC or a compatible machine(see Appendix A for how to copy it).

3. EMG-to-sound problem

3.1 Haskins EMG data

The Haskins EMG data (Baer, Alfonso, and Honda, 1988) available at ATR include the activities of the six extrinsic tongue muscles corresponding to an utterance type, /əpVp/, where V is one of 11 English vowels. Four principal muscles, the genioglossus anterior (**GGa**) and posterior (**GGp**), the hyoglossus (HG), and the styloglossus (SG), are considered, because they participate in the control of the tongue positions (Honda, 1991; Honda, 1992) and can be related to the positional articulatory parameters. In addition, EMG response of the lip muscle, the orbicularis oris superior (**OOs**) that contributes to the lip rounding and protrusion, is used. These EMG responses were obtained by the ensemble average of 10 tokens of rectified and smoothed EMG activities for each utterance. Data resolution is 12 bits

(4096 points) and the sample rate is 5 ms per frame. The response of each muscle is time-aligned at the target vowel onset that corresponds to the EMG frame number 100. The data also include jaw movements in the sagittal plane and are in the same format as the EMG.

The data are accompanied by an audio tape which contains speech signals simultaneously recorded with the EMG and jaw movement data. In order to label the EMG responses, signals of the 11 utterances are visually segmented on their spectrograms and waveform printouts. The segment boundaries are then specified in terms of the EMG frame number aligned at the vowel onset in each utterance. This boundary information is stored in a document file, "DOC.TB". A detail description of the document file is provided in Appendix C.

It should be mentioned here that the acoustic manifestation of the EMG response lags by about 100 ms (*i.e.*, 20 EMG frames). The segment boundaries, therefore, must be shifted by -20 frames on the EMG response display. It is not so clear why such a large delay occurs in the acoustic manifestations; nevertheless in order to align the OOs response peak and the silence just before the intervocalic /p/ release, it is necessary to advance the boundaries by about 20 frames. I assume that EMG responses of all the muscles are advanced by 20 frames in respect to the acoustically defined boundaries.

3.2 combin.c

The purpose of an accessory program, **combin.c**, is to organize the EMG channel files and create single EMG-jaw movement file for each utterance. The other important function is to detect the maximum EMG activity value of each muscle over 11 utterances to be used for the normalization. For the jaw y-position data (vertical movement), the mean value over the 11 utterance is calculated and then this mean value is subtracted from the original data. The absolute maximum value, then, is detected to be used for the normalization of the vertical jaw movement patterns.

The file name format of the original EMG ASCII files is as follows :

CjUk.ACS,

where j is a EMG channel number and **k** a utterance number (from 1 to 11). These files are stored in the directory C:\maeda\emg\data\. The EMG

(and jaw movement) responses are shifted up by a fixed bias, 2048. In **combin**. This bias is subtracted from the original responses. Reading six channels (one vertical jaw movement and plus five EMG responses) for each utterance, **combin** creates the single output file (also in ASCII) in the following file name format :

EMGk.TB

in the same directory as the original channel files. Note here that the utterance number **k** is now specified always by two digits as 01, 02,...., 09, 10, and 11. You may print out these files. The first line is the initial frame number (always 1 in this particular data set) followed by the final frame number (always 200). The number on the third line indicates how many channels are assembled (6 in this case). The fourth line contains the labels identifying the channels, which are immediately followed by the maximum value of each channels. The movement and EMG response data for 200 frames are listed successively after a space line.

3.3 From EMG to vocal tract configurations : **emglam.c**

The hart of **emglam** is a set of linear equations to convert EMG responses (stored in EMGk.TB) into the values of articulatory parameters. Since, how to run this program is described in Appendix B, I shall mainly describe here the idea behind the conversion.

During an informal discussion at ATR, Kusakawa has reported that vowels plotted on an EMG space forms a classical vowel triangle. One axis is HG - GGp (the hyoglossus activity minus the genioglossus posterior activity). The other axis perpendicular to the previous one corresponds to SG - GGa. The paired muscles function antagonistically to each other. Thus HG and GGp would contribute to a back- low/front-high tongue body movement, while SG and GGa to a back-high/front-low movement Honda (1991). But do those movements suffice to specify the acoustic characteristics, specifically F1 and F2,of the vowels? This question leads me to think about the calculation of the formant patterns from the EMG data.

It is interesting to note that tongue movements inferred from the EMG responses correspond well with the effects of specific parameters in the articulatory model. The effects of the tongue body parameter appears as a back-low/front-high movement corresponding to the HG - GGp dimension. Those of the tongue shape parameter results in a bulged/flat

deformation of the tongue body, which roughly coincides with the SG - GGa dimension. I have postulated, therefore, linear relations between the paired EMG responses and the articulatory parameters, **tp** (tongue-body position) and **ts** (tongue-body shape), as follows :

$$tp = c2HG + c3GGp$$

and

$$ts = c4SG + c5GGa,$$

where c2, c3, c4, and c5 are fixed coefficients and their values must be determined empirically (The same notations are used in the **emglam.c** and in **emgfrmnt.c**). The EMG data include the activity of **OOs** that contributes to the lip rounding gesture. The OOs activity, therefore, can be related to **lh** (lip height parameter) as

$$lh = c6OOs,$$

where c6 is a fixed coefficient. At present, OOs determines only **lh**. It should be interesting to vary also the lip protrusion parameter **lp** as a function of OOs response. This modification might improve to some extent the calculated F1 and F2 frequencies of /u/ in the utterance /əpup/ described latter. Note also that without the response of the muscles antagonistic to OOs, **lh** has always a negative or zero value, thus the lip aperture can only decrease from a neutral value depending on the degree of **OOs** activities The absence of the antagonistic muscle activities in the **lh** determination must be considered when interpreting the model calculated lip shapes in **emglam** and formant frequencies in **emgfrmnt**.

How can such linear relationships be justified? The EMG responses represents muscular force and articulatory parameters are essentially positional. The linear (proportional) relationships between the force and position mean that a spring system obeying Hooke's law is assumed as

$$\Delta x = cF,$$

where Δx is a displacement from the equilibrium position (corresponding to the parameter value) and F is a muscular force measured by the EMG response. Thus cN (N = 2, 3, .., 6) in the above equations can be regarded as a "spring" constant. The spring model is particularly convenient, since each linear component in the articulatory model describes the "deviation" from

the neutral (actually mean) vocal tract configuration, that is specified by the corresponding parameter value. In this modeling, each articulatory parameter value is determined uniquely by the responses of a single antagonistic pair of muscles independent of the activity of other muscles.

For jaw position, the normalized vertical jaw position data (**JAW**) are translated into the jaw parameter (**jw**) by a simple proportional principle as

$$jw = c1 JAW.$$

The values of the six coefficients are determined by the following procedure. The EMG responses and jaw position data are normalized by dividing them by the corresponding maximum values which were detected in **combin** as described before. The range of the EMG responses become from 0 to 1. Considering the antagonistic combination of the paired muscles, the maximum range of combined EMG activities becomes from -1 to 1. Since, the maximum value of any articulatory parameters is generally within the range from -3 to 3 standard deviations, it is reasonable to let all initial guess values be +3 or -3. At the initial step therefore, the range of **jw**, **tp**, and **ts** is from -3 to +3, and that of **lh** is from 0 to -3. The coefficient values are then empirically oriented so that a strong constriction would not occur during a vowel segment but does occur during every /p/ closure in the /əpVp/ utterances.

The tentative values of the six coefficients in the current programs, **emglam** and **emgfrmnt**, are listed in Table I.

- Table I - Values of coefficients for the EMG-to-position conversions.

c1	c2	c3	c4	c5	c6
-2.5	3.0	-2.0	3.0	-2.0	-1.5

These coefficients are defined at the beginning of each of the two current programs as global variables and can be easily modified.

Note that in this "spring" formulation, there is no lag between force (EMG responses) and displacement (tract configurations and then acoustic characteristics). It is generally observed that the EMG responses precede the corresponding acoustic pattern of speech by as much as 100 ms as mentioned earlier concerning with OOs. In order to align the temporal

variations of the EMG derived articulatory parameters and utterance segments, the displayed boundaries are shifted toward the left by 20 frames (= 100 ms) as defined by "**EMGadvance**" in **emglam.c**. It has been assumed therefore that the EMG responses of every tongue and lip muscle equally advance from the corresponding movements by 100 ms.

An example of such display is shown in Fig.5 for the utterance, /əpip/.

3.4 From EMG to formant frequencies : **emgfrmnt.c**

This program reads the EMG**k**.TB and calculates the articulatory parameters and then vocal tract configurations as in the case of **emglam**. But it does not plot either parameters or tract profiles. Instead, **emgfrmnt** computes area functions and then the formant frequencies (F1, F2, F3, and F4) of the two vowel segments in each of the 11 utterances. The calculated frame-by-frame variations of articulatory parameters and formants during two vowel segments are written to a file with the following name format :

FRM**k**.TB.

These files are created in the directory, C:\maede\emg\data\. An example of the result of formant calculations is graphically shown in Fig.6 for the same utterance /əpip/ as in Fig.5.

Let me describe some preliminary results of formant calculations to show that the "linear spring" model for EMG-to-position conversions works rather well. The conversion coefficient values listed in Table I have been used.

After calculating temporal formant patterns with **emgfrmnt**, F1 and F2 frequencies of four vowels, /i, æ a, and u/, are sampled at fame number 95. Considering the 20 frame advance of the EMG responses and the vowel onset aligned at frame 100, the sample point corresponds to 15 frames (75 ms) after the onset. The F1 and F2 values are then compared with the measures published by Honda for these four vowels (Honda, 1991).

The calculated and measured F1 and F2 for the vowel /i/ in the utterance "APEEP" (/əpip/) are shown in Table II. The result for the vowel [æ] in the utterance "APAEP" (/əpæp/) is shown in Table III.

Table II Vowel [i]

	F1 (Hz)	F2 (Hz)
measured	350	1841
EMG derived	360	1899

Table III Vowel [æ]

	F1 (Hz)	F2 (Hz)
measured	660	1239
EMG derived	602	1248
open-lips	624	1259

The model calculated F1-F2 pattern with EMG as input compares rather well with the measured pattern. Still, a certain degree of discrepancy occurs between the measured and the calculated values, especially for F1 of [æ]. The calculated one is 58 Hz lower than the measured counterpart. Of course, our objective is not to obtain an exact F1-F2 match. This would be meaningless, since the EMG and speech data are taken from a single speaker while the articulatory model is based on the factor analysis of X-ray data of a different subject having globally different vocal tract size and shapes. A more reasonable criterion for judging the quality of the match is whether or not the EMG derived vowels distribute appropriately on the F1-F2 plane. In this sense, the observed discrepancy in F1 could be accepted, but we shall attempt to gain a better match to demonstrate how articulatory parameter values can be modified in a sensible manner.

Consulting a diagram which displays the effects of a change in individual articulatory parameters on F1-F2 patterns (Maeda, 1991), an effective correction for raising the F1 frequency is to open up the lips. F1 and F2 frequencies are recalculated using vtcalcs by modifying the lip

height parameter value from its original value -0.22 (closed) to 0.0 (a neutral position). The other parameters remain unchanged from their values determined by EMG and by jaw data. The resultant F1 and F2 are indicated on the third row in Table III, labeled "open-lips". Now, the F1 and F2 values better match the measured ones, since F1 is raised by 22 Hz, although there is a slight overshoot (11 Hz) of F2.

Table IV shows the results of the F1-F2 calculation for the vowel [ɑ] in the utterance "APAP" (/əpɑp/).

Table IV Vowel [ɑ]

	F1 (Hz)	F2 (Hz)
measured	658	983
EMG derived	584	1040
open-lips	633	1086

As in the case of the previous vowel [æ], the calculated F1 is too low in comparison with the measured one. The lip aperture is increased by specifying **lh** value equal to 0.2 (more opened) instead of the EMG derived value, -0.21 (rounded). New F1 and F2 values are calculated using **vtcalcs** and the result for open-lips is listed in the bottom row of Table IV. With this manipulation, F1 is raised about 50 Hz, resulting in a reasonable F1-F2 pattern for this vowel.

The results for the vowel [u] in the utterance "APOOP" (/əpup/) are summarized in Table V.

Table V Vowel [u]

	F1 (Hz)	F2 (Hz)
measured	372	927
EMG derived	384	1322
backed-tongue	424	1050
rounded-lips	384	1007

In this case, the EMG derived F2 is much too high. Observation of the model generated vocal tract profile for the same frame number using **emglam** indicates that the tongue body is not backed enough. F1 and F2

values are recalculated using vtcalcs with the tongue-body position parameter set at 2.00, instead of the EMG derived value, 0.72. The results are shown in the third row, labeled "backed-tongue". It is seen that F2 is substantially decreased from 1322 Hz to 1050. The F1 frequency, however, increases from 384 to 424 Hz. In order to lower the F1 and F2 frequencies, lip rounding is increased from the EMG derived value of -0.61 to -.8, resulting in a reasonable F1-F2 pattern for the vowel [u] as shown in the row labeled "rounded-lips". F1, F2 lowering should be also possible by an increase in the lip protrusion.

Although, it was possible to obtain a correct formant pattern for [u], the large correction required for the **tp** value is disturbing. The value of **tp** is defined from EMG responses of the antagonistically paired **HG** and **GGp**. In the cases of the previous two vowels, [æ and ɑ], the improvement has involved only lip parameters in which the correction is justified because of the lack of EMG response of the muscles antagonistic to **OOs**. The correction of the tongue body position is puzzling, since in the utterance "APOPE" (/əpoup/), the formant pattern of the final portion of the diphthong, corresponding to /u/, is correctly predicted from the EMG responses as F1 = 400 Hz and F2 = 900 Hz. In this case, the EMG derived tongue-body position is appropriately backed as expected for this vowel. Perhaps, the speaker has articulated differently, and it might be the case that the articulation of the simple vowel /u/, involved activities of some other muscles which are not measured in the EMG experiments. I shall discuss more on the tongue-body position predicted from the EMG responses in the following section.

4. Epilogue : EMG specified vowel formant patterns

The insufficient tongue backing in [u] predicted from EMG deserves additional investigation. This is a reason for conducting a follow up experiment in Paris after my two month visit at ATR. In the previous section, we have investigated only four selected vowels. If EMG derived formant patterns are compared with measured ones for all the 11 vowels, the comparison might shed some light on the true ability of the EMG responses in the vowel specification.

To obtain a reference F1-F2 scatter for the 11 vowels, F1 and F2 frequencies of each vowel are visually determined from the spectrogram and the spectrum slice of each utterance 75 ms after the vowel onset. The results are shown in Fig.1. The F1-F2 scatter roughly corresponds to the

standard vowel system of American English. The relative positioning of low vowels, [æ, a, and ʌ] does not conform to the typical vowel system. This is, at least in part, due to the fact that I have measured F1 and F2 from a single token of each utterance. It should be noted that the two vowels [e and o] were diphthongized. This deviation from the standard vowel pattern doesn't matter much, however, since the objective here is to compare the acoustically measured F1-F2 frequencies with those calculated from the EMG responses sampled at the same instant.

The result of the F1-F2 scatter computed from the EMG responses is shown in Fig.2. The coefficient values used in EMG-to-articulatory-parameter conversion are the same as before, and are listed in Table I. Comparing the measured F1-F2 scatter (Fig.1) and one derived from the EMG (Fig.2), it should be noticed that there are two major discrepancies. The first discrepancy occurs for the three back vowels, [u, U, and o], whose F2 frequencies are systematically too high. A too high F2 value for a back vowel can be explained by assuming that the EMG derived tongue position is too fronted. The same observation applies to the vowel [ʌ], in which the EMG version is too centralized in comparison with the measured one. The second discrepancy is noticeably low F1 frequencies for the low vowels, [æ, a, and ɔ]. A too low F1 frequency for a low vowel can be explained by assuming that lh determined from OOs alone resulted in excessively small lip aperture, as described before.

It becomes clear that corrections are necessary to obtain a reasonable fit between these two F1-F2 scatters. It can not be corrected in an arbitrary manner however, since the model can produce any combination of F1 and F2 values within the vowel space. It is always possible to match perfectly model derived F1 and F2 against measured ones by individually adjusting the articulatory parameters. To obtain perfect match by correcting each EMG derived parameter value, therefore, is meaningless.

The model parameter corrections, therefore, are meaningful only for testing a hypothesis such as why the F2 frequency of certain back vowels are too high or why the F1 frequency of certain low vowels turns out to be too low. The above assumptions concerning the too high F2 for the back vowels and too low F1 for the low vowels are tested by the following rules for corrections: If F2 is too high for a vowel, place the tongue in a more posterior position or vice versa. If F1 is too low, then widen the lip aperture or vice versa. These two rules imply that we only manipulate the values of two parameters, tp (tongue-body position) and lh (lip opening height).

Table VI Values of lip-tube height parameter, lh

	EMG derived	Modified
[æ]	-0.22	0.1
[a]	-0.21	0.5
[ɔ]	-0.6	-0.3
[u]	-0.62	-0.9

Let us modify the values of lh first. The values of lh derived from EMG and modified, are listed in Table VI. Notice that for the first two vowels, [æ and a] in Table VI the "rounding" indicated by the negative values are changed to positive values corresponding to "unrounding" of the lip opening. In the third vowel, [ɔ], the degree of lip rounding is decreased. The vowel [u], however, is more rounded. This was necessary to compensate for the effect of the tongue position modification to be described below. The tongue backing resulted in too high F1 frequency for [u]. Thus, it was necessary to round the lips more in order to offset the raised F1 frequency.

The results of lh correction for the four vowels are indicated by the open circles in Fig.3. Arrows connecting the original(closed circle) and the modified (open circle) indicate the effect of lh correction. For the three vowels, [æ a and ɔ], both F1 and F2 frequencies raise as expected. But, while the degree of raising due to the increase in lip aperture is greater for F1 than F2 in the case of [æ and a],. the effect of enlarged lip opening for [ɔ] is primarily F2 raising. In the case of [u], the lip rounding lowers F1 and F2 frequencies as expected. But F2 is still too high for [u] The F1-F2 scatter seems to be more regular, but the F2 frequencies of the back vowels are systematically too high.

In Table VII, the EMG derived values of tp parameter and their modified values depending on vowels are listed.

*Table VII Values of tongue-body position parameter, **tp***

	EMG derived	Modified
[I]	1.37	-0.6
[ʌ]	1.92	3.0
[ɔ]	1.9	3.5
[o]	1.09	2.0
[U]	1.47	3.0
[u]	0.73	2.5

The results of tp correction for the six vowels are plotted with open circles in Fig.3. Again, arrows indicate the shift between original and adjusted values. Note that both lh and tp are corrected for the two vowels [ɔ] and [u]. The scatter now corresponds relatively well to the measured one shown in Fig.1. Thus, the model calculation confirms the hypothesis that the EMG data lack two important pieces of information; one for lip "spreading" and the other for tongue backing.

There is one more thing to consider; namely the EMG derived tongue-body position. It might be suggested that correct tongue-body positions can be obtained without parameter modifications by optimizing the EMG-to-position conversion coefficients, specifically c2 associated with HG and c3 with GGp. This will not work for following two reasons. First, if the value of c2 was increased in order to enhance the tongue backing function of HG, then this would result in the complete closure of the vocal tract during certain vowels. Second, as can be seen in Table VII, the tongue must be fronted for [I] contrary to the other vowels. In other words, the correction of tp is selective depending on the identity of vowels. The value of c2 in particular cannot be modified effectively for all vowels.

The important consequence of this is that activity of the antagonistic pair of the extrinsic tongue muscles, HG and GGp, is not sufficient to determine the front/back dimension. It is speculated that other extrinsic muscles, such as sternohyoid, participate in vowel articulation and that their activity is also necessary to specify vowels.

There is other evidence which favours this conclusion about the present EMG data. There are cases where HG responses before the p-release cannot simply be translated into tongue backing movements. (This issue was raised by Fujimura during the informal conference by Kusakawa at ATR.) An example of temporal patterns of articulatory parameters derived from EMG is shown in Fig.5 for the utterance /əpip/. The cursor (the vertical solid line in the figure) points to the vowel offset at frame number 50. The corresponding vocal tract profile and the frontal lip shape is show in the left-top corner. The profile clearly exhibits a backed tongue body position. This backing of the tongue is puzzling, because such a movement is not needed for production of the target vowel [i], which is a front-high vowel. Observing the temporal pattern for the tongue-body parameter (the second curve from the top), the tongue is maximally backed during the p-closure (frames 51 to 60) and maximally fronted at frame 90 during the [i] onset. This temporal variation is the direct consequence of the HG response indicated by the thin solid line in Fig.5. Now the question is whether the tongue of the speaker was actually backed during the p-closure as indicated by the HG response.

The tongue movement data are not available. However, the question can be verified indirectly by comparing EMG derived formant patterns and the spectrogram of the recorded speech signal. The EMG derived formant patterns are shown in Fig.6. It is observed that the F2 contour falls toward the offset of the initial schwa vowel which is the direct consequence of the tongue backing movement. A similar F2 falling pattern can be seen toward the offset of the target vowel [i] also. A spectrogram of the utterance /əpip/ is shown in Fig.7, which indicates no particular falling F2 pattern at the initial vowel offset. The F2 falling at the offset of [i] in the spectrogram is much weaker than that in the EMG derived F2. It is therefore reasonable to conclude that the speaker's tongue is not actually backed, which is contrary to what is expected from the HG response. This finding implies that the HG and GGp responses alone cannot specify completely the vowels.

5. Concluding remarks

This preliminary study has shown that the Haskins EMG data contain surprisingly rich information for deriving vowel patterns from EMG, even though the data seem to lack information needed to specify the tongue front/back dimension. It is noted, however, that our linear formulation for converting the EMG responses to the articulatory positions might be too simplistic. Possibly, the relation between force and displacement is not

linear. Moreover, articulatory positioning specified by the activity of the paired antagonistic muscles might not be independent from that of other muscles, in opposition to our assumption. Obviously the proposed EMG-to-sound formulation must undergo a further improvement. The value of this study, perhaps, is in the demonstration of how effectively the EMG-to-sound approach can provide an objective means to evaluate the activity of the individual muscles in terms of the articulatory positioning and acoustic patterning, thus filling the gaps between the EMG activities and the acoustics.

Acknowledgment

I thank Kiyoshi Honda and Eric Vatikiotis-Bateson for insightful discussion and comments on this report. Eric's proof reading has made this report more readable. I also wish to thank Yoh'ichi Toh'kura for inviting me to work at the ATR Human Information Processing Laboratories.

References

Baer, T., Alfonso, P.J. & Honda, K. (1988) Electromyography of the tongue muscles during vowels in /ə pVp/ environment. *Annual Bulletin of Research Institute of Logopedics Phoniatrics*, University of Tokyo, **22**, 7 - 19.

Honda, K. (1991) *a manuscript for* "A statistical analysis of tongue muscle EMG and vowel formant frequencies," *Journal of Acoustical Society of America*, **90**, 4 (Part2), 2310 (abstract).

Honda, K., Kusakawa, N. & Kakita Y. (1992) An EMG analysis of sequential control cycles of articulatory activity during /əpVp/ utterances. *Journal of Phonetics* **20**, 53-63.

Honda K. (1992) Physiological background of speech production (tutorial). *Journal of Acoustical Society of Japan*, **48** (1), 9 - 14 (in Japanese).

Kakita, Y. & Fujimura, O. (1984) Mapping from muscular contraction patterns to formant patterns in vowel space. *Transaction of the Committee on Speech Research, The Acoustical Society of Japan*, **S83-100** (March 31, 1984) In Japanese

Kakita, Y., Fujimura, O. & Honda K. (1985) Computation of mapping from muscular contraction to formant patterns in vowel space. In *Phonetic Linguistics*, (V.A. Fromkin, editor), pp 133-144, Academic Press Inc.

Kusakawa, N., Honda, K. & Kakita, Y. (a forthcoming paper).

Maeda, S. (1979) Un modèle articulatoire de la langue avec composantes linéaires. In *10émes JEP, GALF*, pp. 152 - 164.

Maeda, S. (1982) A digital simulation method of vocal-tract system. *Speech Communication*, 1, 199 - 229.

Maeda, S. (1990) Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In *Speech Production and Speech Modeling* (W.J. Hardcastle & A. Marchal, editors), pp. 131 - 149. Kluwer Academic Publishers.

Overall, J.E. (1962) Orthogonal factors and uncorrelated factor scores. *Psychological Reports*, 10, 651 - 662.

Figure captions

Figure 1 F1-F2 scatter of 11 American English vowels measured on spectrograms.

Figure 2 F1-F2 scatter of the 11 vowels calculated from the EMG responses.

Figure 3 EMG derived F1-F2 scatter. For the vowels indicated by the open circles, the values of lip height parameter (lh) are corrected as shown in Table VI. Each arrows indicates the change from the original position to that after the correction.

Figure 4 EMG derived F1-F2 scatter. For the vowels indicated by the open circles, the values of tongue-body position parameter (tp) are corrected (in addition to the lh corrections) as shown in Table VII.

Figure 5 EMG derived frame-by-frame variation of 6 articulatory parameters along the utterance "APEEP" (/əpip/) at the left. The vocal tract profile and the frontal lip shape (at the left-top) corresponds to the frame number 50. The parameter variations are indicated by the thick lines. The tp variation is derived as the weighted sum of HG (indicated by the thin line) and GGp responses (indicated by the dotted line). Similarly, the ts variation is derived from SG (the thin line) and GGa (the dotted line) responses.

Figure 6 EMG derived formant patterns (from F1 to F4) along the utterance "APEEP" (/əpip/). Formant frequencies are calculated from the 6 articulatory parameters shown in Fig.5. The "*" indicates the lip closure period before the intervocalic p-release.

Figure 7 Spectrogram and waveform of the recorded signal corresponding to the utterance "APEEP" (/əpip/).

Appendix A

xcopy and compiling programs

The software package is stored on two floppy disks, and can be transferred to the hard disk on **C:** using MS-DOS command, **xcopy**. Start copying from disk1 by

C:\>**xcopy a: c: /s/e**

The second disk contains the remaining files, so do

C:\>**xcopy a:\maeda\emg c:\maeda\emg /s/e**

Normally, this will work and you are ready to run user ready demonstration programs, **lam**, **vtcalcs**, and **emglam**, as explained in Appendix B. However, if you want to modify a program and compile it, you have to change the first instruction line in the file, C:\maeda\myinc**mydir.h** as follows :

```
From    char *BGIdir = "C:\\maeda\\grafix\\BGIdir\\";
to      char *BGIdir = "C:\\TC\\";
```

assuming that you have installed Turbo C (TC) in the directory C:\\TC. This modification is necessary to properly plot the graphics, although I don't understand why it is necessary.

Appendix B

Documentation for **lam**, **vtcalcs**, and **emglam**

These three programs are written with a standard C language using Borland Turbo C v2.0 (TC). Graphics functions in the programs extensively uses TC graphics primitives which are not standard. Thus it is most convenient to modify them using TC.

To run programs, change directory by entering

C:\>cd maeda

There are three *.BAT files to run each of the three programs, in the directory "maeda". By typing out these .BAT files, you will see in which sub-directory these executable and the source programs are located.

lam

This program demonstrates effects of each of the seven articulatory parameters upon the vocal tract profile and the frontal lip shapes. Any VT configuration is described by the linear combination of the effects of the seven parameters.

After entering

C:\maeda\>lam

you will see seven articulatory parameter codes, **jw**, **tp**, etc.. Enter one of the seven codes. Now a vocal tract profile and an elliptic lip frontal contour appear on the screen. To stop the motion, hit any key, for example a space bar. To see the effect of another parameter, hit the space bar again. But if you want quit, hit the letter "q". You are going back to the directory "maeda".

The parameter codes mean

jw : jaw position (affects on shapes of the lips, tongue, and larynx).

tp : tongue-body position
ts : tongue-shape
tt : tongue tip position

lh : lip height (aperture)
lp : lip protrusion

lx : larynx height.

vtcalcs

This program calculates acoustical characteristics of a stationary vocal tract, as the transfer function, formant frequencies, and vowel signal. Physical parameters, sound velocity, air density, etc., and area-function specification are modified by menu operations.

After entering "vtcalcs" as

C:\maeda\>vtcalcs

you will see a root menu. A menu item is "pointed" by using the **arrow keys** on the keyboard. To select the item, hit the enter(CR) key. Or you can hit the first letter of the menu label, to point and select that item. You will see a sub-menu or a prompt for entering a new value. To go back to the previous menu, hit the "ESC" key.

This is a "what you see is what you get" program, so you just play with the arrow, CR, and ESC keys.

When you want to calculate the transfer function and formant frequencies, or to synthesize a stationary vowel **from an arbitrary area function**, (with menu "from file"), you must create area-function files beforehand. The files should be in the directory, C:\maeda\vtcal\data, by default. But you can create them in any directory. In this case, you must modify the "Area directory" in the menu item "Save option" of the root menu. It is easy to figure out the file format (in ASCII) by reading an existing area file, for example, c:\maeda\vtcal\area\iy.ARE. By the way, area-function files have a fixed extension, ".ARE".

A stationary vowel with a fixed F0 contour (about 300 ms in duration) is synthesized when you select the menu item "**Synthesize**" using a specified area function. It will take a relatively long time to finish the synthesis. You must patiently wait for the calculation to be finished. The synthesized vowel and calculated glottal wave forms are stored in temporary files, respectively "tempsig.SIG" and "tempglt.SIG" in the directory "C:\maeda\sig\". The file names and the directory can be changed by

selecting "Save options". If you want to save the vowel signal, hit "k" (for "keep"). The program creates the save file, "VOWELj.SIG", where j is an integer and is incremented every time you select "Keep". The glottal signal is not saved. The file identifier "VOWEL" can be also modified using "Save options". As you may notice, signal files have the fixed extension, ".SIG". By the way, the vowel signal has 16 bit resolution and a fixed 10 kHz sampling rate.

emglam

This program demonstrates the variations of EMG derived articulatory parameters and the corresponding model calculated vocal-tract configurations along syllables, /əpVp/ where V is one of 11 American English vowels. After entering

C:\maeda\>emglam

you will see the display for the first utterance, /əpip/ ("APEEP"), the articulatory variation at the left and the tract profile and frontal lip shape at the right. The vocal tract corresponds to the frame pointed by the mouse cursor. The six articulatory parameters displayed are, from top to bottom, jaw position, tongue-body position, tongue shape, tongue tip, lip height and tube length (protrusion). The larynx height parameter is not included here. Also notice that the parameter values of the tongue tip and lip-tube length are kept zero, since EMG data are not available for deriving the values of these parameters. The unit for all articulatory parameters is the standard deviation. The value of the tongue-body parameter is derived by the weighted sum of HG (in red) and GGp (in green) and of the tongue shape by that of SG (in red) and GGa (in green). The value for the lip height parameter is simply the (inverted) weighted OOs EMG activity, since the EMG data of its antagonistic muscle is not available.

To proceed to the next utterance, push the **right mouse button**. To abort the program at any utterance, hit the **left mouse button**.

Appendix C

EMG data document file, **DOC.TB**

This text file contains segmental information on 11 utterances, /əpVp/, used for the EMG data collection. I have segmented manually each of the 11 utterances by observing its spectrogram. Only the tokens in the first series of the 10 repetitions are used for the segmentation. The corresponding averaged EMG data are sampled at the rate of 5 ms per frame, and the onset of each target vowel, "V", corresponds to EMG frame number 100. It is possible to define and align the segment boundaries in terms of the EMG data frame numbers for each utterance. The boundaries are used to relate the temporal patterns of EMG responses and converted articulatory parameters with the phonetic sequence of the utterance in **emglam** and **emgfrmant**.

In addition, the file contains the utterance number in two digits, and the initial frame number (always 1 in the data), final frame number (always 200), *etc.*, which are used by **combin**. The initial part of the file is listed below :

```
11

01  1  200  5
52  70  91  100  154  200
A  *  P  EE  P

02  1  200  5
51  68  91  100  154  200
A  *  P  I  P
.
.
.
```

Note that the "11" on the first line is the total number of utterances. Then spaced by one empty line, the second line indicates the utterance number, the initial frame number, the final frame number and the number of segments, "5" in this case. In the following line, 6 numbers indicate the segment boundaries in terms of the EMG data frames. The last line is the 5 segment labels. The star, "*" indicates the silence before p-release.

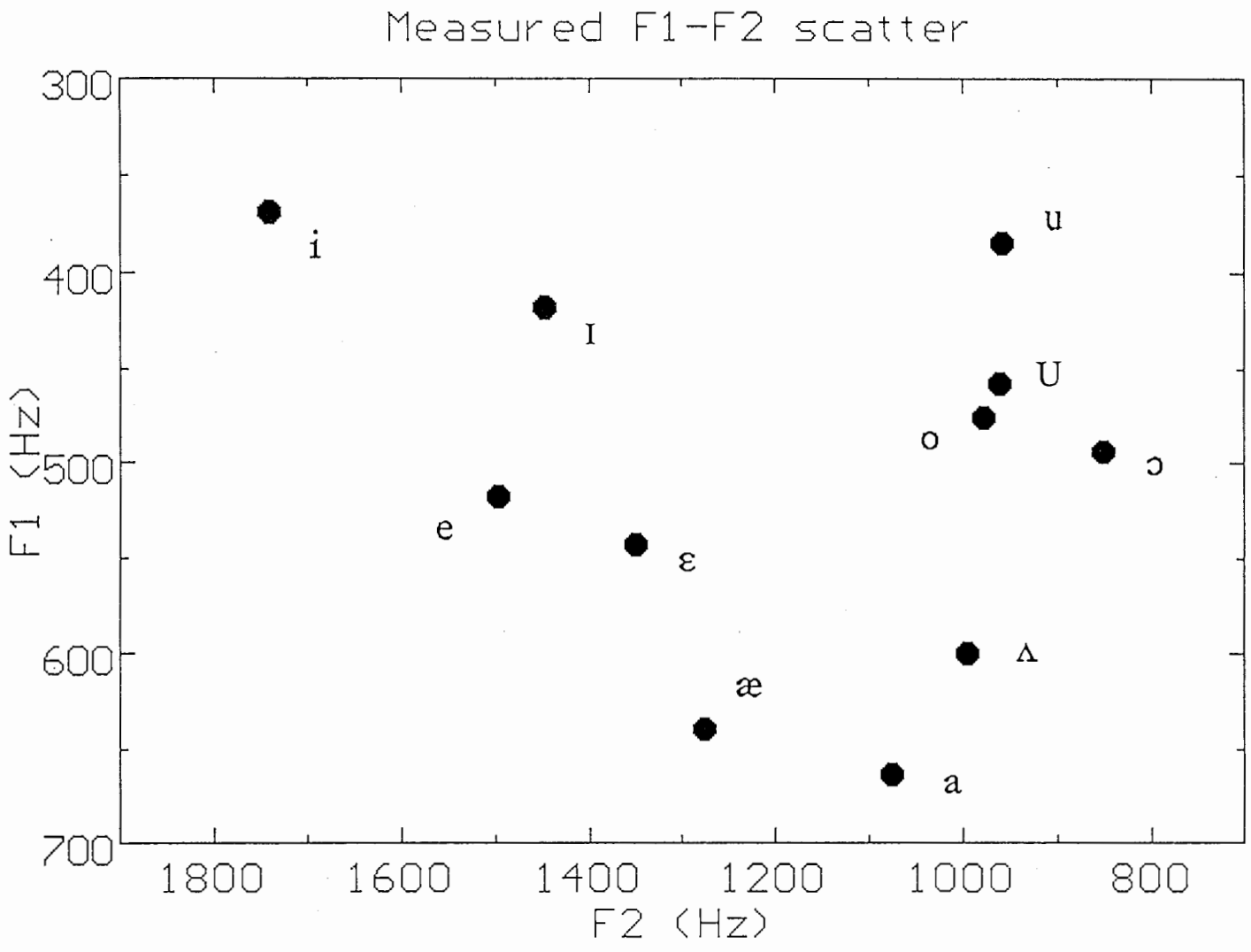
Measured F1-F2 scatter
300
400
500
600
700
F1 (Hz)
1800
1600
1400
1200
1000
800
F2 (Hz)
i
I
e
ɛ
æ
a
ʌ
u
U
o
ɔ

Fig. 1

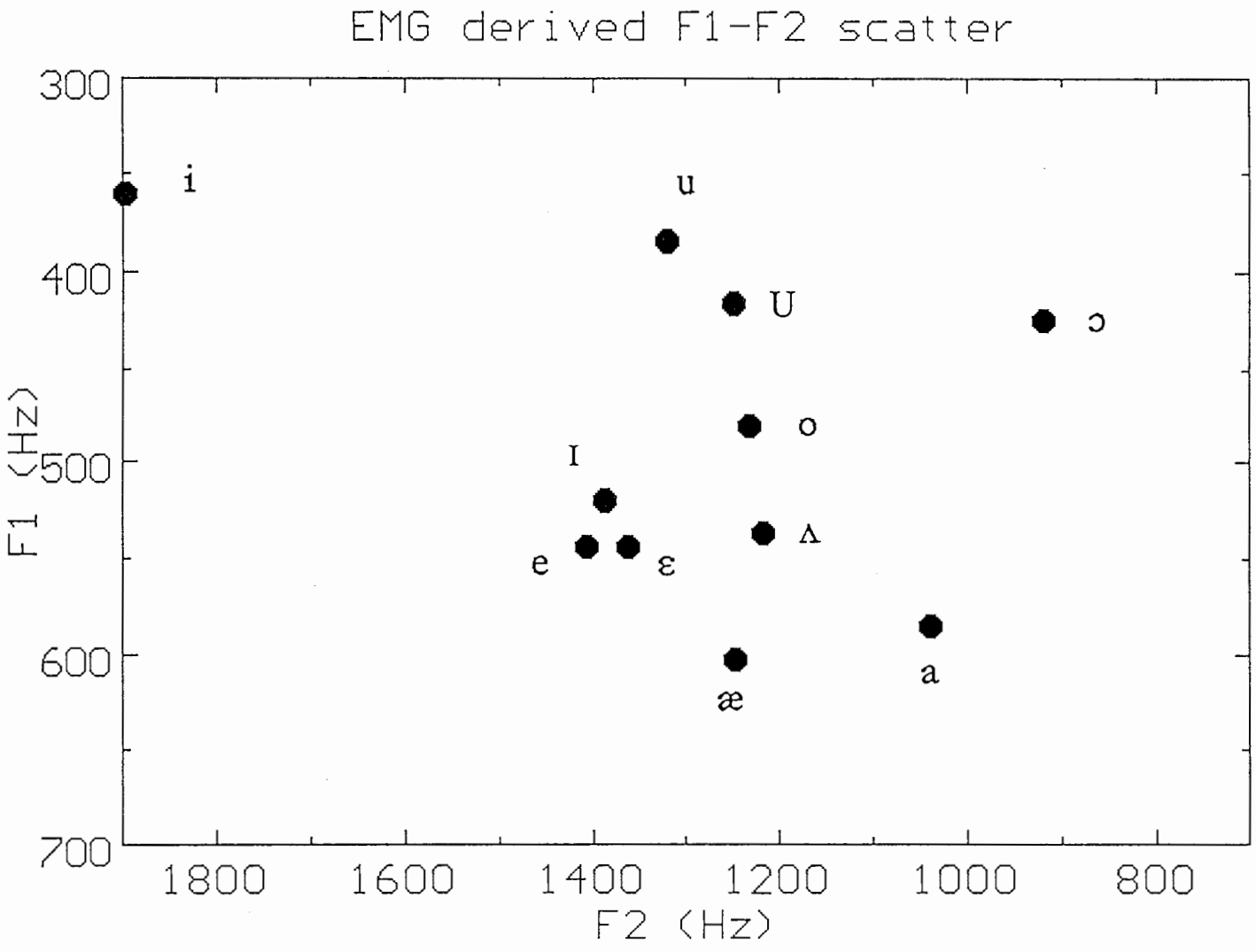
EMG derived F1-F2 scatter
300
400
500
600
700
F1 (Hz)
1800
1600
1400
1200
1000
800
F2 (Hz)
i
u
U
ɔ
o
I
ʌ
e
ɛ
æ
a

Fig. 2

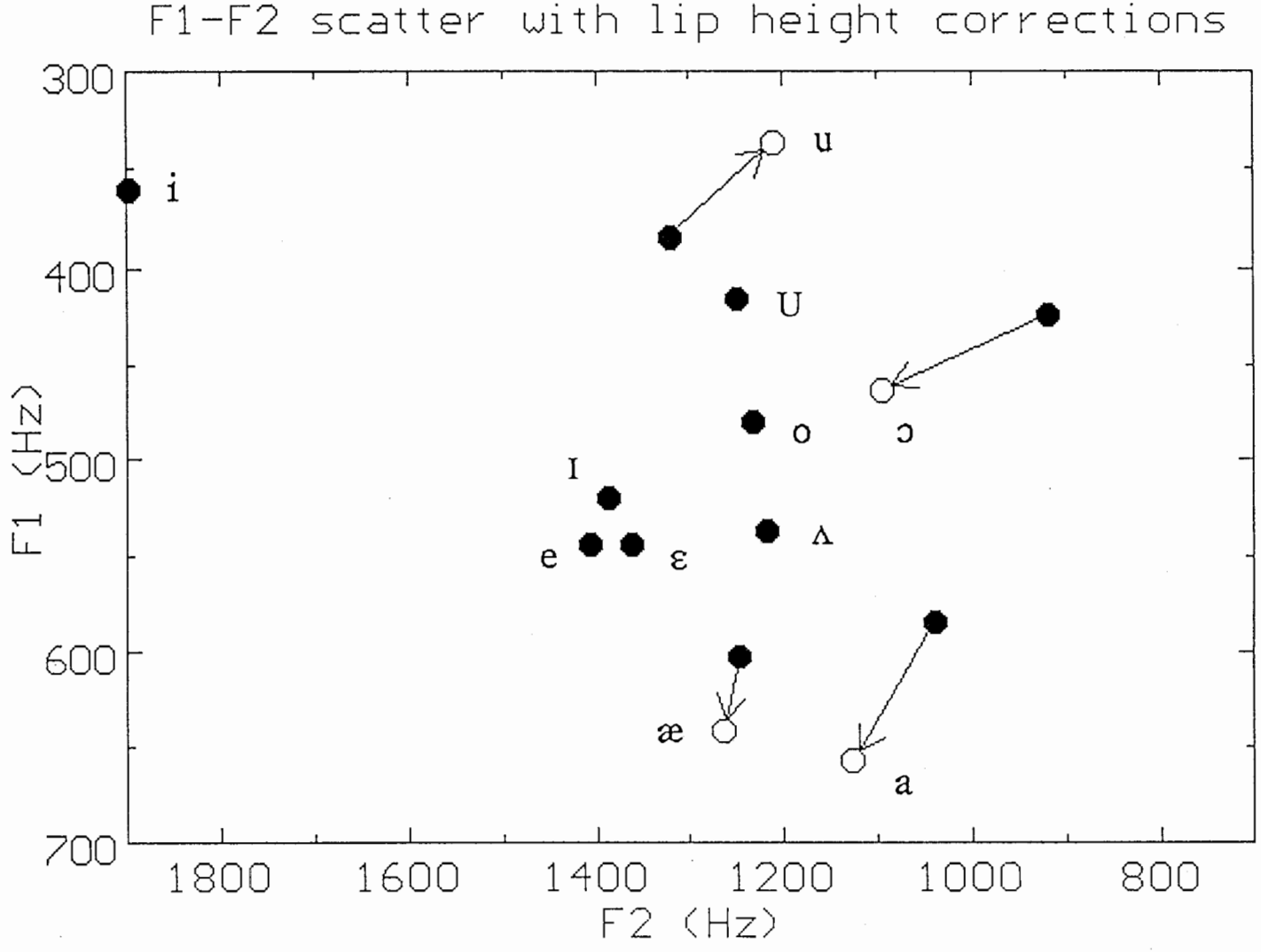
F1-F2 scatter with lip height corrections
300
400
500
600
700
F1 (Hz)
1800
1600
1400
1200
1000
800
F2 (Hz)
i
u
U
o
ɔ
I
e
ɛ
ʌ
æ
a

Fig. 3

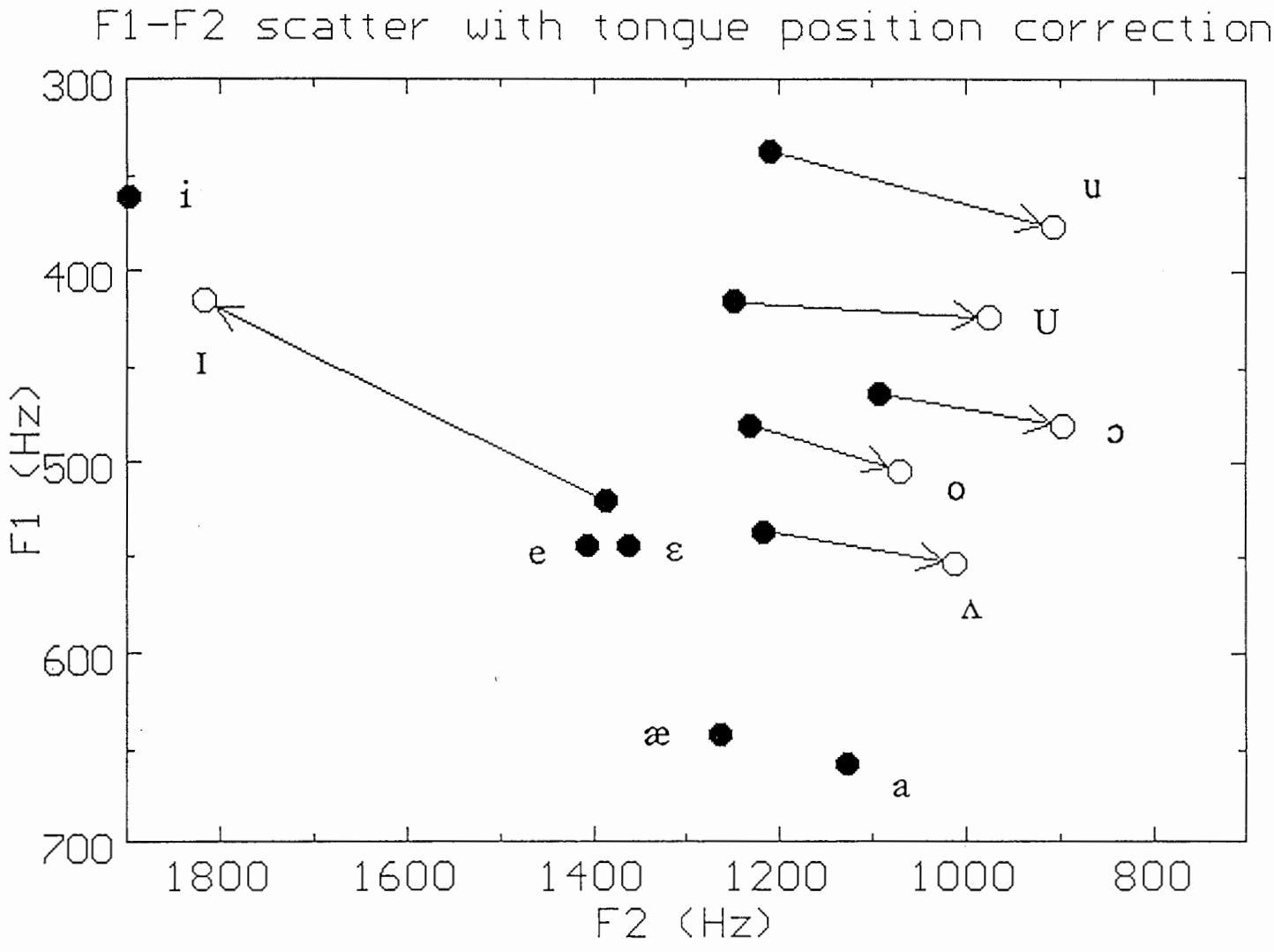
F1-F2 scatter with tongue position correction
300
400
500
600
700
F1 (Hz)
1800
1600
1400
1200
1000
800
F2 (Hz)
i
I
e
ɛ
æ
a
u
U
ɔ
o
ʌ

Fig. 4

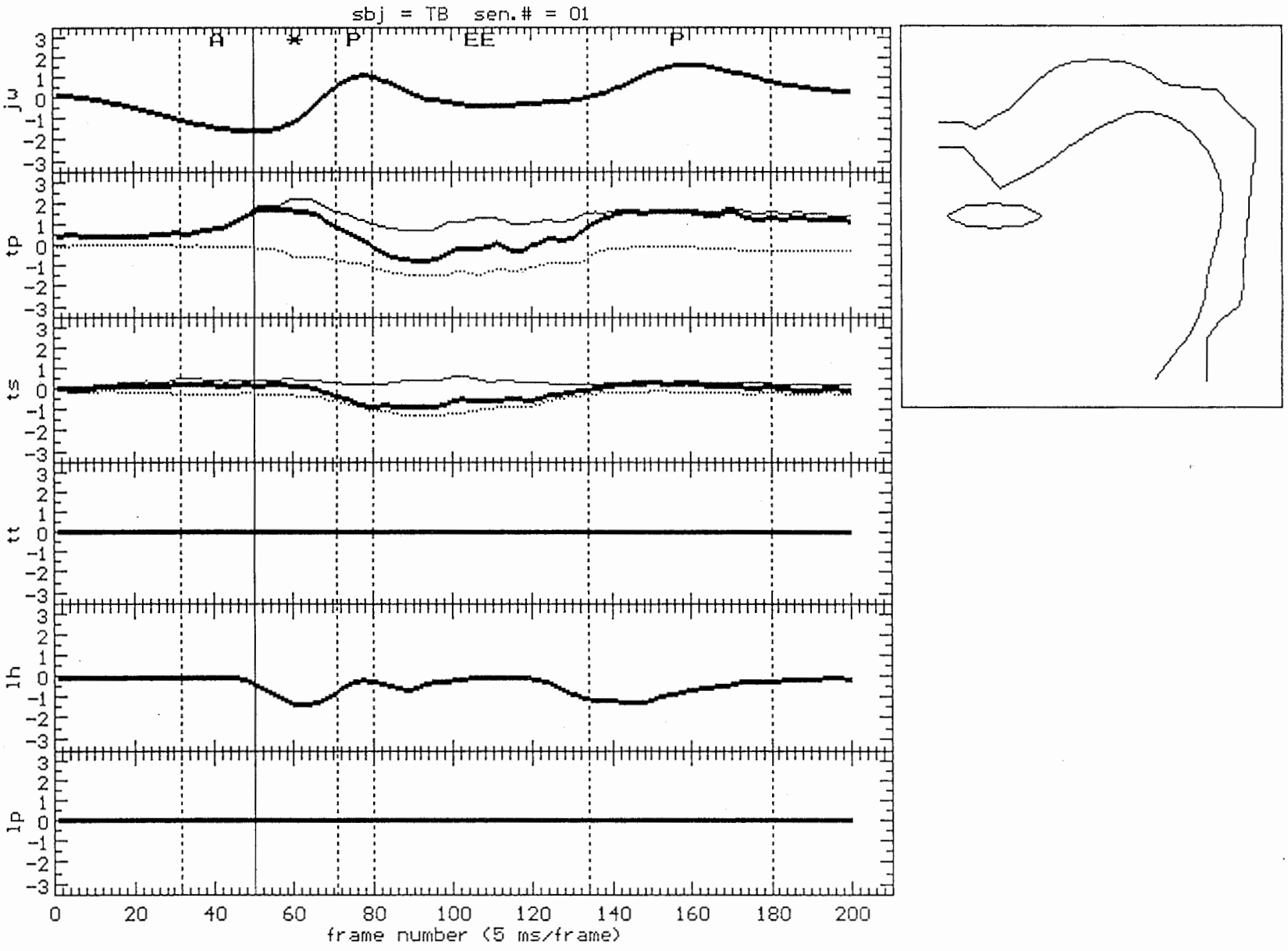
sbj = TB sen.# = 01
A
*
P
EE
P
jw
tp
ts
tt
lh
lp
0
20
40
60
80
100
120
140
160
180
200
frame number (5 ms/frame)

Fig. 5

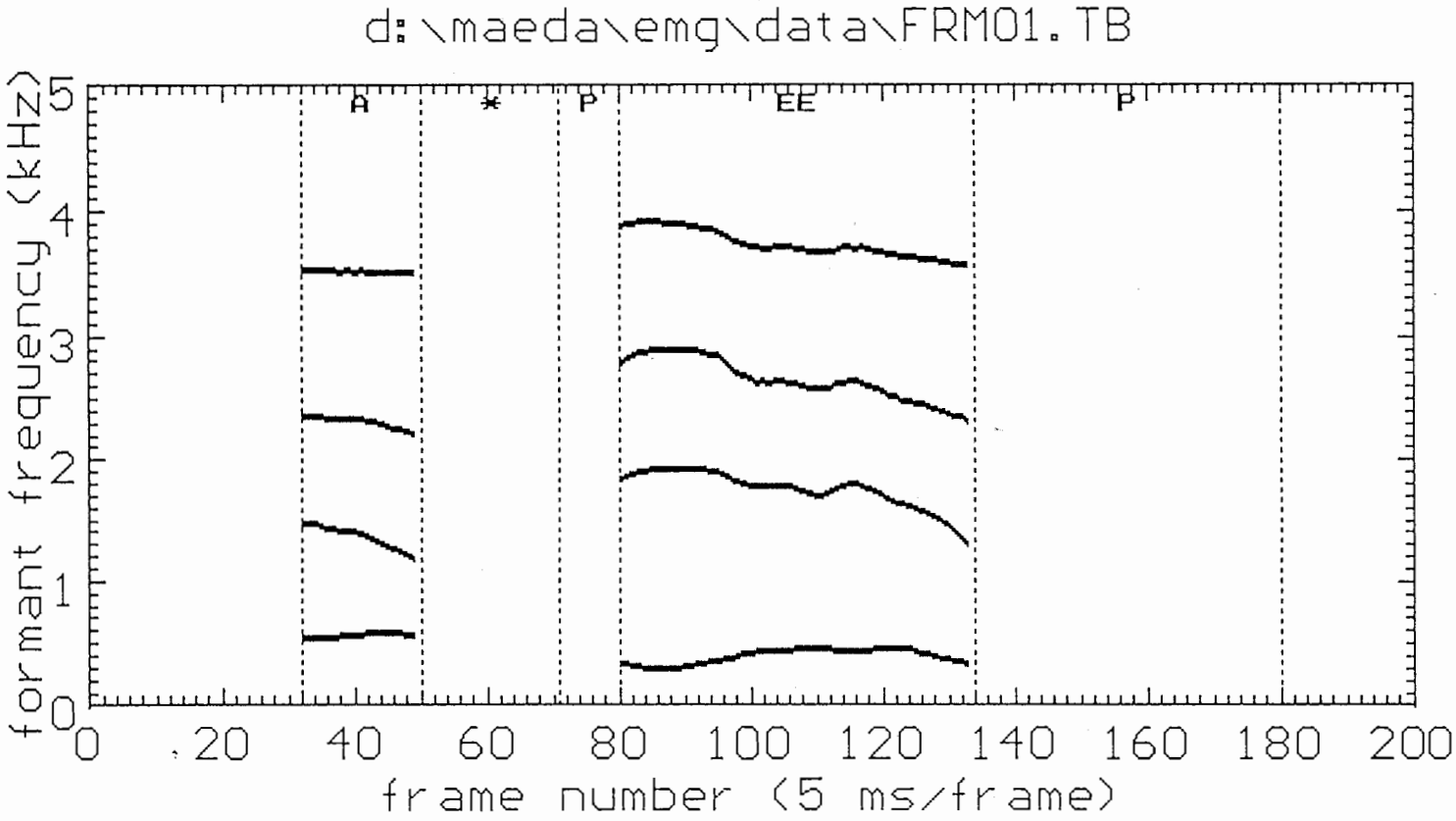
d:\maeda\emg\data\FRM01.TB
formant frequency (kHz)
0
1
2
3
4
5
A
*
P
EE
P
0
20
40
60
80
100
120
140
160
180
200
frame number (5 ms/frame)

Fig. 6

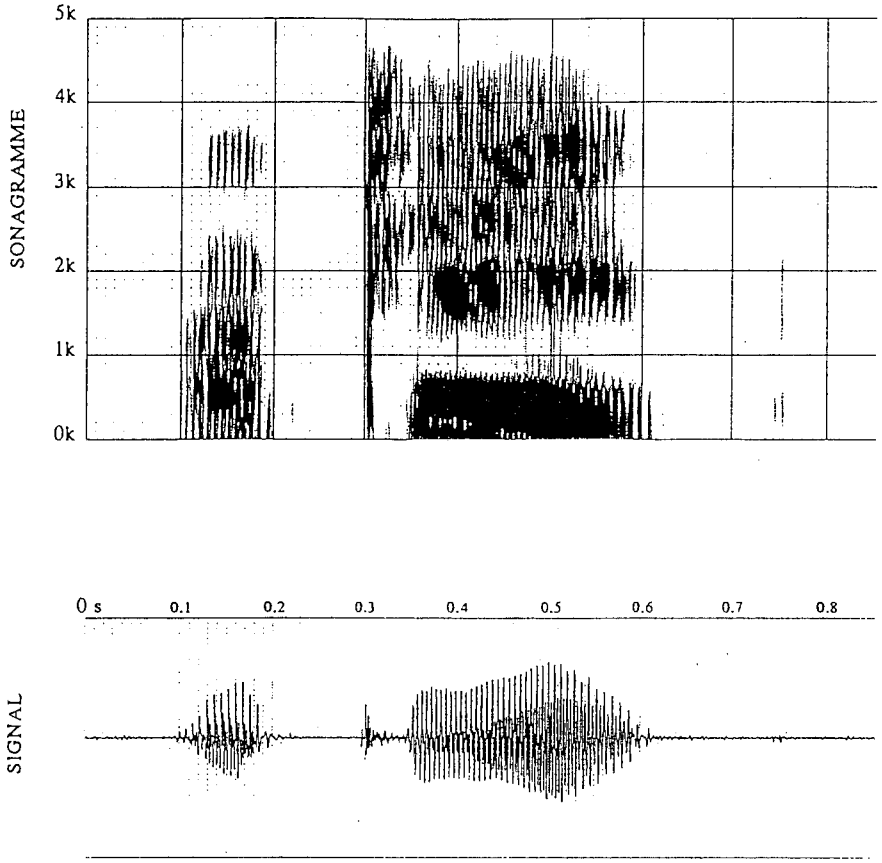

Fig. 7