Ann.Bull.RILP(1989)No.23, 27-37

EFFECTS OF SPEAKING RATE ON FORMANT TRAJECTORIES AND INTER-SPEAKER VARIATIONS*

Satoshi Imaizumi and Shigeru Kiritani

Introduction

In order to improve the quality of synthetic speech, rules for generating the temporal patterns of formant trajectories under various speaking rates must be constructed. There are still numerous differences among the conclusions of studies on the effects of speaking rate[1-11]. Some studies[2,3], based on formant value measurements, indicate that increased rates of speech result in systematic deviations in the obtained formant values from their putative targets, that is, "vowel reduction". Some others[4,5,6] claim that such "vowel reduction" does not always occur at fast speaking rates. Furthermore, other studies based on observation of the articulatory movements, claim that adjustments in speaking rate are achieved by strategies which differ among speakers[8,9], and on the carefulness of their articulation[10]. According to some electromyographic investigations, control of the speaking rate is achieved through a reorganization of motor commands[11,12]

As for the relation between speaking rate and coarticulation, research results show large discrepancies, and it cannot be denied that they differ widely according to the speaker, or the speaking style. In order to rationalize this problem, it is necessary to take into account variation factors with clear-cut characteristics, and to find out which characteristics of the dependent variables are more invariable and which ones vary.

In this study, we have investigated the invariable and variable characteristics of vowel formant trajectories with the following variation factors: embedding in a consonantal context, speaker and speaking rate within speakers.

Method

Subjects

Three adult male speakers of the Tokyo dialect with no speech (or sensory-motor) abnormality (N_1, N_2, and N_3) and two adult Broca's aphasics with apraxia of speech (P_1:female and P_2:male) participated in the experiment. For the sake of testing the analysis methods, we also used the speech of one adult man and one adult woman. The aphasic subjects had normal articulatory and phonatory organs and were potentially able to execute any

*The content of this paper was presented at the 2nd Joint Meeting of the Acoutical Societies of America and Japan(14-18 November 1988, Honolulu, Hawaii), J. Acoust. Soc. America, Supple 1, 84, NN18, S128, Fall 1988.

articulatory gesture. Their apraxia was due to a defective control at the central level. We used these aphasic subjects in order to describe the influence of a defective control on formant trajectories. More details are given in our previous report[13].

Since subject P_1 also showed a slight hoarseness, we finally decided not to include her in the results.

Speech Material

The speech material used here consisted of $/V_1CV_2CV_1/$ trisyllables, with C beibg /b/ or /g/ and V_1, V_2 being /a/ or /i/, embedded in the frame sentence /korewa --- desu/ (This is ---.). The formant frequencies of V_2 were to be analyzed. In addition, in order to check the analysis method, sentences consisting of vowels and semi-vowels were recorded.

All the subjects were recorded at two different speaking rates, slow (S) and fast (F) for the normal subjects, normal (N) and fast (F) for the patients with apraxia of speech. Subjects were instructed to avoid unnaturalness and to articulate clearly.

The normal subject N_1 was also instructed, on another day, to speak as clearly as possible at a slow rate (C) and as quickly as possible (Q).

Each subject determined his/her fast or slow (normal) rate. The subjects uttered each sentence five times.

Measurements

Various measurements were synchronized with the speech waveform: images of the tongue mid-sagital section by means of ultrasonic tomography[14], intra-oral pressure by means of a miniature pressure sensor, and vocal fold vibration waveform indirectly obtained by means of electroglottography (EGG).

Images of the tongue mid-sagital section were video-recorded with a 60 frames per second rate. The image frame synchronization pulse, speech waveform and other analog signals were simultaneously recorded with a PCM data recorder.

The intra-oral pressure was recorded to detect the closure interval of the plosive consonants. The EGG signal yielded glottal closure intervals which were used for the pitch-synchronous covariance LPC analysis. Images of the tongue mid-sagital sections were recorded for future use.

Speech Analysis

First, we used an autocorrelation linear prediction analysis, with a fixed frame length of 25.6 ms, for estimating the formant trajectories.

In order to make the formant frequencies and tongue mid-sagital sections correspond to each other, the tongue image frame synchronization pulse was used to set the analysis frame rate. However, because of the requirement of the autocorrelation LPC analysis based formant tracking method, the frame shift was set to the fourth image frame pulse period, that is about 4.2 ms.

In order to obtain more exact values for the formant frequencies in the unsteady portions of the speech signal, we used a covariance LPC analysis, with pitch synchronous frames corresponding to the glottal closure intervals, derived from the EGG signal[15]. We call this method a "closed-phase covariance LPC analysis". Glottal closure intervals were derived from the EGG signal in the following way: Childers et. al. reported that the glottis closure times correspond to the positive peaks of the EGG signal time derivative, and the opening time roughly corresponds to the negative peaks[15]. Thus, each analysis frame in our study could be determined so as to begin at one positive peak and end at the following negative peak, the legth being $T_a(n)$, as shown in Fig. 1. However, there is always a lag between the EGG signal and the speech signal. For this reason, the beginning of the analysis frame was shifted about 1ms later. It's length, $T_a(n)$ was left unchanged.

To summarize, we extracted F_0 by a peak picking method applied to the EGG time derivative, the speech signal energy curve (power), and in the closure portion of voiced plosive consonants we performed an all-pole model LPC analysis.

Finally, in order to derive formant trajectory production rules, we focused our attention on the formant frequencies at the voice onset, on the rate of formant transition at the voice onset and on the formant frequencies at the center of the vowels.

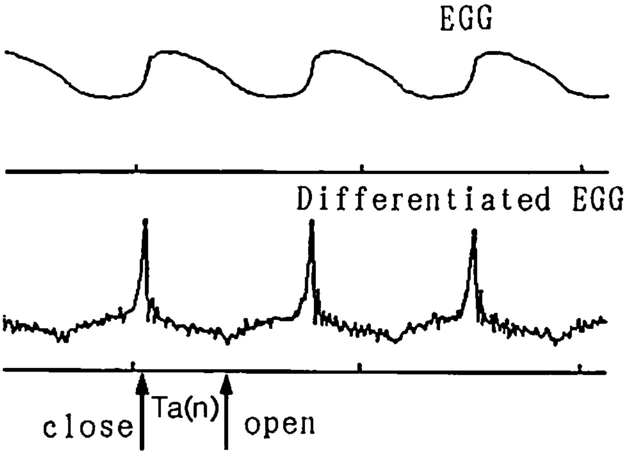
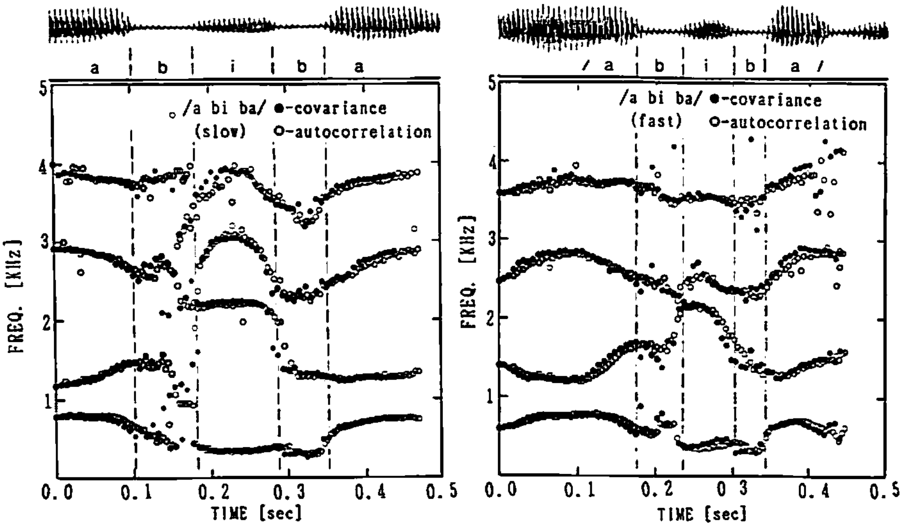


Fig. 1. Estimation of the closed phase of the vocal cord vibration based on the EGG signal time derivative.

Results

Comparison of the Two Methods of Analysis

Figure 2 shows the formant trajectories obtained by using the two analysis methods for the trisyllable /abiba/ pronounced at a slow and fast rate by speaker N_1. The " ● " correspond to the "closed phase covariance LPC analysis", whereas the " O "



(a) (b)

Fig. 2. Formant trajectories for a slow utterance (a) and a fast utterance (b) of /korewa abiba desu/ obtained by the autocorrelation LPC analysis (O) and the closed phase covariance LPC analysis (●).

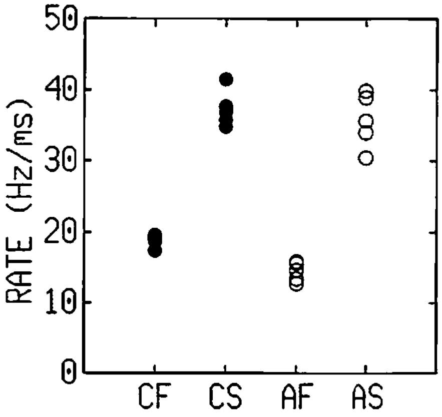


Fig. 3. Absolute value of the rate of second formant transition from /i/ to /b/ in /abiba/. C: Closed-phase covariance LPC; A: Autocorrelation LPC; F: Fast and clearly; S: Slow and clearly

correspond to the autocorrelation LPC analysis synchronized with the tongue image frame pulses. The two methods yielded somewhat different results, especially in the vowel to consonant, or consonant to vowel, parts. In paticular, in the transition from /i/ to /b/ in /abiba/, the pitch synchronous closed-phase covariance LPC analysis could follow the fast transition of the second formant, but the other analysis could not.

The second formant transition rate from /i/ to /b/ was computed based on a least square error approximation by a straight line. The results are shown as absolute values in Figure 3. It appears that the closed-phase LPC analysis yields a larger velocity for the two speaking rates than that obtained by the autocorrelation LPC analysis.

Effects of Speaking Rate on Formant Trajectories

If we compare the formant trajectories in Fig. 2(a) and Fig. 2(b), it appears that when the speaking rate increases, F_2, F_3, and F_4 are lowered. This tendency particularly holds for F_3. The extension of frequency lowering for F_2 was estimated as being in the central portion of the vowel /i/, where as F_1 was approximately flat during roughly 30ms by means of linear regression. The results for the five repetitions are shown superimposed in Fig. 4, where the linear regression line is displayed in 25ms intervals. The formant values differ from one speaker to anoth-

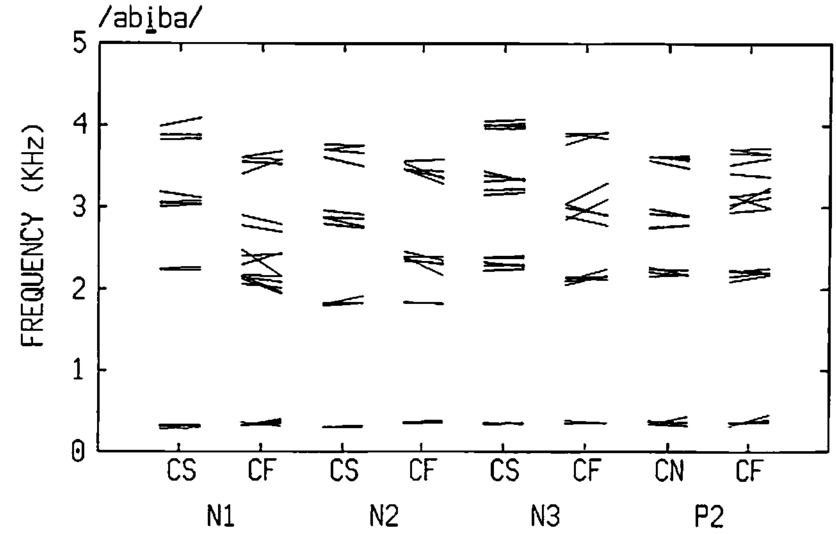


Fig. 4. A linear approximation of the formant trajectories at the center of /i/ in /abiba/ for four subjects N1, N2, N3 and P2. The first C of the pairs on the abscissa indicates the results obtained by the closed phase covariance LPC analysis. The second letter of each indicates the instruction given to the speakers; S:slow and clearly; F:fast and clearly; N:normal and clearly.

er, but the speaking rate has a similar effect in all the normal subjects: F_1 and F_2 are not much affected, whereas F_3 is lowered to a large extent.

As for the aphasic subject P_2, fast articulation only affects F_3, which is a little heightened. So, the speaking rate exerts opposite effects on normals and this aphasic subject with apraxia of speech.

Effects of Instruction

Figure 5 shows the formant trajectories in the central part of the vowel /i/ in /abiba/ for the speaker N_1 with four kinds of instructions. The first letter 'C' of the symbols on the abscissa indicate the closed-phase covariance LPC analysis method which measured these. The second letter indicates the kind of instruction given to speaker. 'C' for "slow, as clearly as possible"; 'S' for "slow and clearly"; 'F' for "fast and clearly"; and 'Q' for "as quickly as possible".

F_1 and F_2 did not change between the conditions 'CC' and 'CS'. Between the conditions 'CS' and 'CF', F_3 and F_4 drop greatly. Between 'CF' and 'CQ', F_2 and F_3 also decrease a lot. Thus, it seems that when the subject was instructed to speak

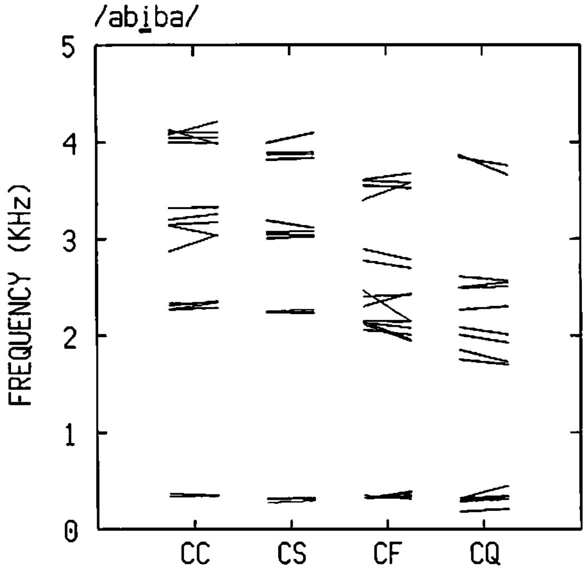
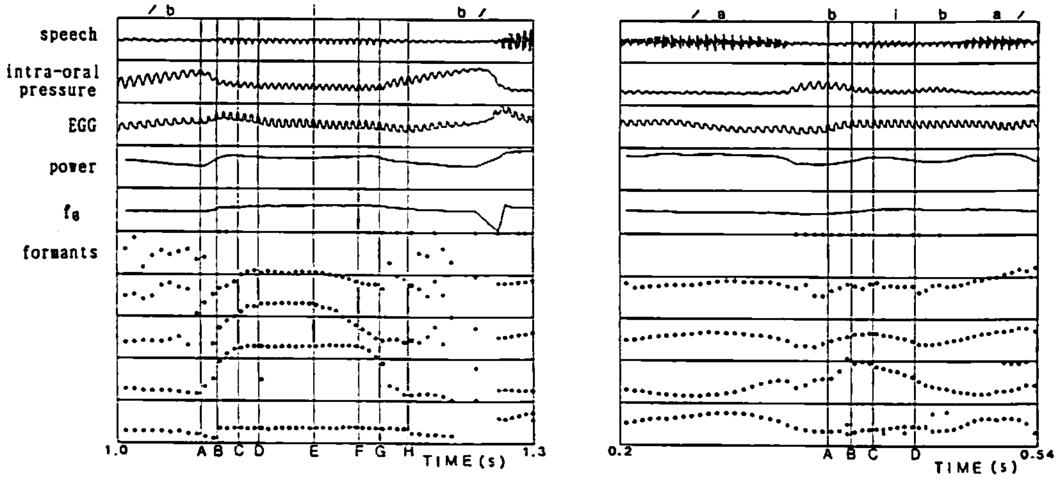


Fig. 5. The linear approximation of the formant trajectories at the center of /i/ of /abiba/ uttered by N1 under four instructions. The first C of the pairs on the abscissa indicates the results obtained by the closed phase covariance LPC analysis. The second letter of each indicates the instruction given to the speakers; S:slow and clearly; F:fast and clearly; Q: as quickly as possible.

"clearly", little vowel reduction occurred with respect to F_1 and F_2, whereas F_3 and F_4 were greatly lowered.

When the subject was instructed to speak "as quickly as possible", vowel reduction appeared for both F_2 and F_3. Furthermore, the variability between repetitions became large.

Figure 6 (a) shows one example of the analysis result taken from the 'CC' condition: "slow and as clearly as possible". Fig. 6 (b) shows one from the 'CQ' condition: "as quickly as possible". As shown in these figures, the differences between the two conditions 'CC' and 'CQ' are quite large not only for formant frequency values, but also for the relationships between the onset or offset timing in each formant transition. In the example of 'CC', the onsets and offsets occur with a certain time lag between the formants. Whereas, in the example of 'CQ', the onset or offset timing of the formant tansitions seem to occur simultaneously. Furthermore, in Figure 6(a), the intra-oral pressure increases largely in the closure portion of /b/ and then begins to decrease rapidly at the release. On the other hand, in Figure 6(b), the increase in intra-oral pressure corresponding to the /b/ closures is much more moderate.



(a) (b)

Fig. 6. One example of /abiba/ uttered "as clearly as possible" (a), and one uttered "as quickly as possible" (b). Both were uttered by normal speaker N1. The symbols A, B etc. on the abscissa and corresponding vertical bars indicate various timings of the onsets or offsets of formant transitions and those for intra-oral pressure.

Discussion

Comparison of the Two Methods of Analysis

With the additional data from the sentences consisting of vowels and semivowels, we can observe the following tendencies.

1) In most of the cases, both methods yield almost identical results in the central portion of vowels, that is, where formants have the steadiest values. However, there are cases where the closed-phase covariance LPC method is able to follow very abrupt transitions that may occur even in the vocalic nucleus, as shown, for instance, for the third formant of /i/ in Fig. 2(b).

2) In transition portions from a vowel to a consonant or from a consonant to a vowel, the two methods yield different results. The closed-phase covariance LPC method can track rapid formant transitions. For example, the two methods will give different results as for vowel initial and final formant values, as well for estimated rates of formant transition in such portions.

3) When the fundamental frequency was high, formant tracking using the closed-phase covarince LPC method was unstable (for example, this occured with speaker P1, for whom the closure portion estimated from the EGG was too short).

To summarize, the closed-phase covariance LPC method was efficient in such cases where abrupt formant frequency changes occurred in transitions at the beginning or the end of vowels by using a short analysis window adopted to the speaker's specificity. However, when F_0 values were very high, or when the estimated closure portion $T_a(n)$ was very short, the closed-phase covariance LPC method was unstable.

Effects of Speaking Rate and Instructions Given to the Speakers.

Figure 4 shows the effect of speaking rate on the formant frequencies in the central portion of the vowel /i/ in /abiba/. For the normal speakers, a fast speaking rate induced a lowering of the F_2, F_3 and F_4 frequencies, and especially of the F_3 frequency. This tendency was common to all the normal speakers investigated.

On the other hand, when the speaker was given the instruction "as quickly as possible" without specifying "clearly", there occurred a large vowel reduction in the F_2 dimension, and the variability across repetitions became large. These results show that the instructions given to the speakers, like "clearly" or "fast", did affect the way they were planning to speak, that is, the way they controlled their articulation, which will affect the characteristics of speech according their perceptual importance. The characteristics, such as F_3, that are of less perceptual importance are possiblely affected to a larger extent.

As for the comparison of the formant transitions at slow and fast speaking rates, it appears that the rate of formant transition is generally faster at a fast speaking rate. However, in the case of /abiba/ shown in Figures 2(a) and (b), the rate of transition was slower at a fast speaking rate.

In Fig. 6, the /i/-/b/ transition for F_3 corresponds to the portion [E,F] of Fig. 6(a) at a slow rate, and to the portion [C,D] of Fig. 6(b) at a fast rate. The transition rate was also slower at a fast rate. One possible explanation of this phenomenon is that when the degree of vowel reduction is large, the extension of the articulator movements becomes small and thus, they cannot reach a high velocity.

When the speaking rate increases, the place of articulation, or constriction of /i/ may move backward to a certain extent, e.g. under the influence of the surrounding back vowels in /abiba/, so the F_2 and F_3 frequencies should be shifted downward and upward, in opposite directions, if the extent of the constriction remains the same16). This is contrary to what is shown in Figures 4 and 5, where F_2 and F_3 are both shifted downward.

It follows that the vowel reduction observed in /i/ in /abiba/ does not solely result from a backward shift in the place of articulation, but rather from a change in the degree of constriction.

As for the influence of the speaking rate on the timing of formant tansitions, whose onsets and offsets may occur simultaneously or with a certain time lag, we are planning to investigate image data for the articulator movements.

The aphasic subject with apraxia of speech, P_2, when instructed to speak "fast", produced speech with a higher F_3 in contrast to normal speakers, as shown in Fig. 4. For other words as well, when instructed to speak at a "fast rate" or at a "slow rate", this subject was highly variable across repetitions.

Conclusions

The present study has obtained the following results.

1) For normal subjects, speaking rate affects vowel formant trajectories in a systematic way. However, the nature and the amount of the changes that are observed also depend on the speaker and on the attention he pays to the instructions he has been given.

2) When speakers are instructed to speak "clearly", whether at a fast or at a slow rate, the formant F_1 and F_2 frequencies of vocalic nucleus of /i/ in /abiba/ are very little affected by the speaking rate. However, F_3 and F_4, whose perceptual importance is relatively small, do vary a lot. With the instruction to

speak "as fast as possible", vowel reduction occurs and lowers F_2 to a large extent.

3) For Broca's aphasics with apraxia of speech, the effects of speaking rate and/or the instruction given are different from those for normal subjects, and the variability across repetitions is large.

These results indicate that those characteristics that are of a lesser perceptual importance are physically affected to a larger extent by changes in speaking rate, and may support the claim that changes in formant frequencies result from the active planning of motor commands.

Acknowledgments

This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas, the Ministry of Education, Science and Culture, Japan. We would like to thank Dr. P. Halle, CNRS, France, for his valuable help in the preparation of this manuscript.

References

1) Miller, J. L.: Effects of speaking rate on segmental distinc tions. in Perspectives on the study of speech. P.D. Eimas and J.L. Miller (Eds.), Lawrence Erlbaum Associates, New Jersey, 39-74, 1981.
2) Lindblom, B.: Spectrographic study of vowel reduction. J. Acoust. Soc. America, 35(11), 1773-1781, 1963.
3) Gay, T.: Effect of speaking rate on diphthong formant movements. J. Acoust. Soc. America, 44, 1570-1573, 1968.
4) Rerbrugge, R. R., and D. Shankweiler: Prosodic information for vowel identity. J. Acoust. Soc. America, 61, S39, 1977.
5) Gay, T.: Effect of speaking rate on vowel formant movements. J. Acoust. Soc. America, 63(1), 223-230, 1978.
6) O'Shaughnessy, D.: The effects of speaking rate on formant transitions in French synthesis-by-rule. Proc. 1986 IEEE-IECEJ-ASJ, Tokyo, 2027-2030, 1986.
7) Miller, J., and T. Baer:Some effects of speaking rate on the production of /b/ and /w/. J. Acoust. Soc. America. 73(5), 1751-1755, 1983.
8) Kuehn, D. P., and K. L. Moll: A cineradiographic study of VC and CV articulatory velocities. J. Phonetics, 4, 303-320, 1976.
9) Sonoda, Y.: Effect of speaking rate on articulatory dynamics and motor event. J. Phonetics, 15, 145-156, 1987.
10) Flege, J. E.:Effects of speaking rate on tongue position and velocity of movement in vowel production. J. Acoust. Soc. America, 84(3), 901-916,1988.
11) Harris, K.: Mechanisms of duration change. (in Speech Communication 2, G. Fant Ed., Almqvist & Wiksell), 299-305, 1974.
12) Gay, T., T. Ushijima, H. Hirose, and F. Cooper: Effect of

speaking rate on labial consonant-vowel articulation. J. Phonetics, 2, 47-63, 1974.

13) Terawa, A., S. Imaizumi, S. Kiritani and H. Hirose:An acoustic study on the speech production mechanisms of Broca's aphasics. IEICE Technical Report (SP88-2), 9-16, 1988.

14) Niimi, S., S. Kiritani and H. Hirose: Ultrasonic observation of the tongue with reference to palatal configuration. Ann. Bull. RILP, 19, 21-27, 1985.

15) Childers, D. and J. Larar: Electroglottography for laryngeal function assessment and speech analysis. IEEE Trans. BME-31, 12, 807-817, 1984.

16) Badin, P. and L. Boe: Vocal tract vocalic nomograms: Acoustic considerations, A crucial problem: Formant convergence. Proc. XIth ICPhS, Se 35.4.2, 352-355, 1987.