Estimating Vocal Tract Area Functions: A Progress Report

F. Lonchamp, J.P. Zerling, J.P. Lefèvre
Nancy, Strasbourg, Le Mesnil St-Denis, France

At the previous International Congress of Phonetic Sciences, Fant remarked that 'very little original data on area functions had accumulated. The 1960 Russian vowels have almost been overexploited'. A quantitative estimate of the variation in the cross-sectional area of the vocal tract is necessary to relate in a cogent way articulatory and acoustic data. Articulatory synthesis which promises high-quality elocution, requires accurate knowledge of area functions. The dearth of cross-dimensional, as opposed to sagittal, measurements is related to difficulties in using photographic data and to the potential hazards of tomographic X-ray exposure. Also, with any method, it is difficult to assess the accuracy of the recovered areas, as we lack normative values. We believe that several techniques should be used concurrently to provide a measure of consistency. We report preliminary results for one subject using two different procedures.

1. Area function determination from the tract impulse response

This first technique was originally devised by Sondhi and Gopinath (1971). The area function is determined from measurements of the vocal tract response to an impulsive acoustic pressure wave. The experimental set-up is shown in fig. 1. Following work mainly at the Electrical Engineering Dpt. of Laval University at Quebec, where the experimental data for this part of the study were gathered, area functions have been published for all French vowels, except the heavily labialized [u] and [y] (Tousignant et al. 1979; Lefèvre et al. 1981, 1983). The procedure is summarized in fig. 2 with numbers illustrating the following steps:

- 1. The acoustic impulse input signal e(t) and its response s(t) after reflection in the vocal tract are both sampled.
- 2. From the delay and magnitude of s(t), the major reflection coefficient (and hence the area value) is located and estimated.
- 3. Due to the bandlimited nature of the signals, smoothing of the area discontinuity with a transition is necessary.
- 4. Using the transmission line model, the tract response is obtained through convolution of the input signal with the impulse response of the vocal tract computed from the reflection coefficients.
- 5. Finally, the residual signal obtained by subtracting the calculated

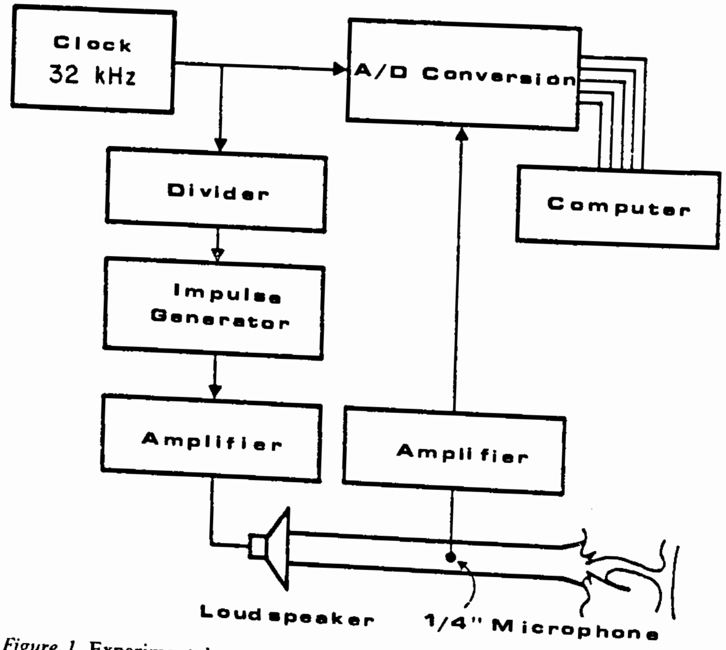

Figure 1. Experimental set-up for impulse response measurements.

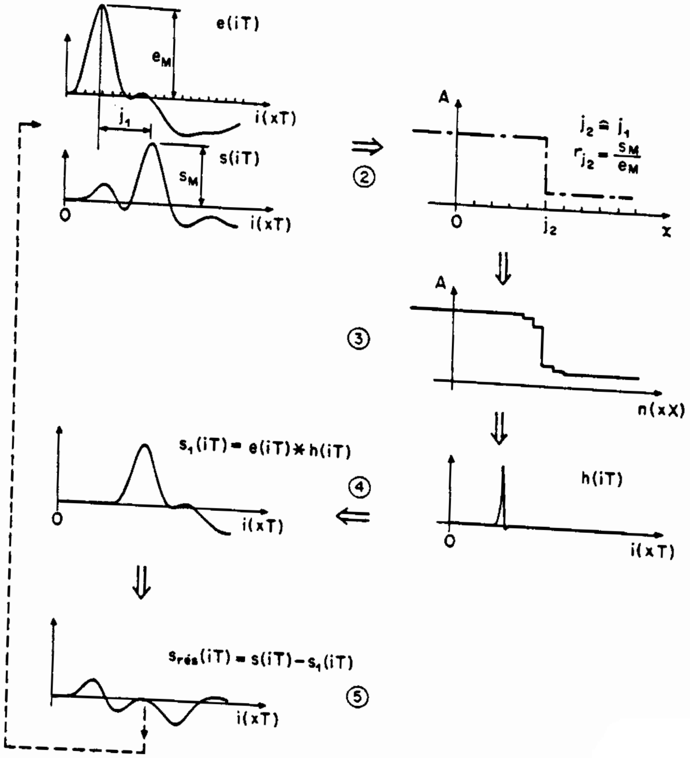

Figure 2. Area reconstruction algorithm.

response from the measured one can be further analyzed by repeating the procedure from step (2).

The area function is thus determined in 10 to 20 passes.

The known advantages of this method (Tousignant et al. 1979; Sondhi 1979) include absolute area estimates and recovery of true length. Drawbacks do exist however: exertion of suitable muscular tension; auditory feedback not available for the subject to check on his articulatory posture. Lip and jaw positions may not be natural while insuring airtight closure at the mouthpiece. To try to control for these last two effects, and to provide for a new test of accuracy, the subject was instructed to phonate as soon as the response measurements had been taken. All results reported here show a good correspondence between formant frequencies computed from the recovered area functions and measured from the tape recording of the session. In other cases, discrepancies were noted. But it is impossible to tell whether the recovered functions are incorrect or whether the subject changed his articulatory posture between measurement and phonation times. Although plausible area functions have been reported for 'constricted' vowels (i.e. [i] or [o]) for another subject with a seemingly larger and longer vocal tract (Lefèvre et al. 1983), only 'open' vowels such as [ε] or [œ] yielded reasonable area functions (cf. Sondhi and Resnick 1983). In all cases, total length and the location of the cavities seemed correct; only their volumes were not. Results for an [ε] token is shown in fig. 3. The dotted line refers to the area function for another [ε] by the same subject, derived from the second technique to be described below. Although discrepancies occur at the glottis and in the mouth region, the general shape is similar. Length values are within 0.4 cm. While the tokens are different, the formant frequencies are close (see caption). Measured and computed frequencies match satisfactorily. The area function is

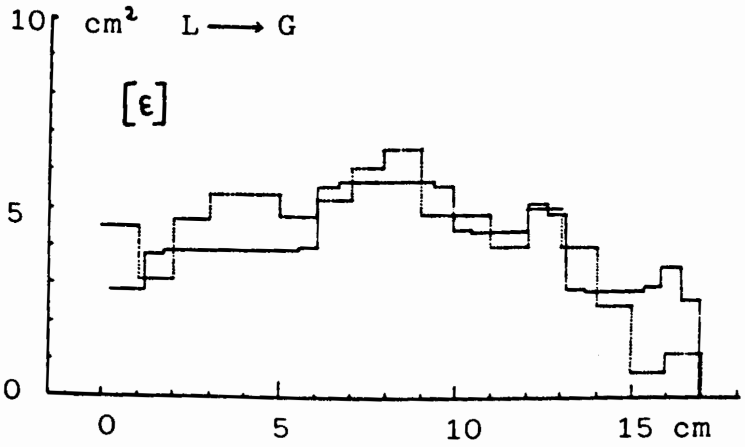

Figure 3. – Area functions for 2 [ε] tokens ___ acoustic impulse response : measured formants: 550-1670-2450-3430; computed formants: 545-1630-2500-3450; · · ·: sagittal to area conversion; measured formants: 580-1650-2512-3520; computed formants: 585-1650-2620-3360.

rather different from the one reported in Lefèvre et al. Fig. 4 illustrates 4 [œ] tokens with close spectral contents. In this case, the area function derived from the X-ray data for another token, shown in the insert, looks very different. The areas are [ɔ]-like in shape. Phonetic description of French has interestingly pointed out the acoustic similarity of [œ] and [ɔ] sounds. We cannot rule out the use of two different tract configurations for the same vowel sound. Clearly, simultaneous X-ray and impulse measurements would have to be made.

2. Conversion of X-ray sagittal measures to area functions

In this second approach, we sought to optimize the numerical coefficients of a set of functions relating vocal tract sagittal dimensions to area values, by simultaneously minimizing the discrepancy between measured and calculated formant frequencies on a set of vowels. A similar approach has been reported by Maeda (1971). Length and sagittal widths at 1 cm intervals were taken from an X-ray film for the vowels [i, ε, a, ɔ u] in an [əb-b] context (Zerling 1979). The first 4 formant frequencies were computed through autocorrelation LPC from the synchronous sound recording. The teeth, uvula hump and epiglottis were ignored. Being more distinct, the midline groove for [ε] and the side tongue outlines for [i] were traced in the upper pharynx zone. A crucial choice is the form and number of functions relating sagittal to area values. As reviewed in Wood (1982), most authors favour a power function for the mouth. The pharynx and larynx regions are modelled usually as a number of ellipses, the cross-dimensions of which (c_i) are set to a constant value. After a long series of pilot experiments, we selected 3 power functions (A (area in cm^2) = $a_i .x^{bi}$; x: sagittal width in cm) for the mouth, uvula and upper pharynx, and the lower pharynx (epiglottic region). Only the laryngeal tube was modelled as an ellipse. As the width of the pharynx never exceeded 2.7 cm, Wood's 'cosine' function was not used. Formant frequencies were computed from the estimated area functions using the transmission line approach (Liljencrants and Fant, 1975). Radiation losses were modelled as an extra section at the lips. Using recent simulation data by Maeda (1982), we were able to express Fw as a function of uncorrected F_1 values (F_1u) in the well-known correction formula ($F_1 = \sqrt{F\omega^2 + F_1u^2}$) for wall vibration effects: $F\omega = 0.04zF_1u + 187$. As no simple formula is available, no correction was made for the effect of open glottis. A random search procedure was used to simultaneously optimize the 7 coefficients for the 5 vowels. As [i, ε, a] gave consistently closer matches, they were more heavily weighed in the decision metric. As a final step, the first (lip and/or teeth) section was slightly adjusted to further improve the fit. The coefficients, error (%F_i) and absolute mismatch values in Hz (Δ F_i) are given below:
- mouth: a=1.95; b=1.53 - uvula and upper pharynx: a=3.00; b=1.40 - lower pharynx: a=2.40; b=1.23 – larynx: c=1.6.

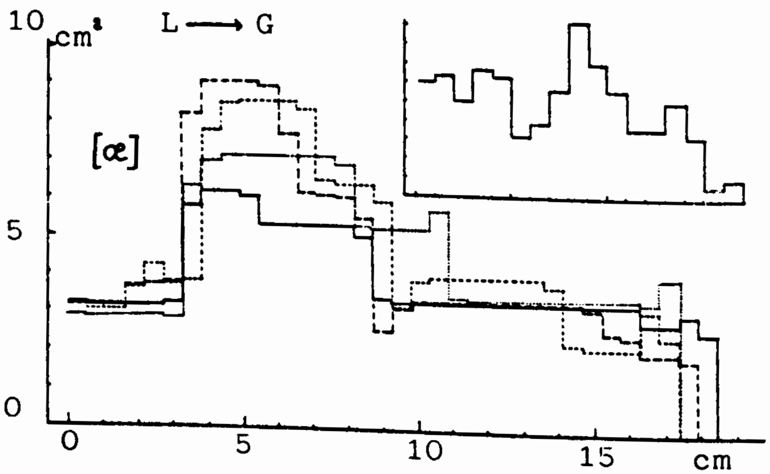

Figure 4. Area function for 4 [œ] tokens. Insert shows area from X-ray data. Measured and computed formant frequencies: —— 560-1275-2425-3310, 545-1296-2426-3330; — — —: 535-1230-2620-3425, 540-1220-2520-3395; ----: 525-1355-2420-3670, 535-1340-2500-3650; . . .: 530-1345-2590-3450, 530-1360-2530-3300.

	%F_1	ΔF_1	%F_2	ΔF_2	%F_3	ΔF_3	%F_4	ΔF_4
[i]	–3.0	10	–0.9	19	–1.5	44	–1.3	45
[ε]	0.5	3	0.0	1	4.5	96	–4.6	163
[a]	–0.5	3	2.0	26	–0.9	22	1.3	45
[ɔ]	4.9	29	4.7	52	–8.1	211	1.4	47
[u]	–4.9	18	8.1	70	–7.0	177	–0.7	25

Figs. 5 and 6 show the area functions for the 5 vowels. The mouth coefficients are almost identical to the values reported by several authors (see Wood, 1982). They also check with values derived from a plaster cast made on the subject. The upper pharynx values are slightly above those for 'level 1' reported by Gauffin and Sundberg (1978) especially for large sagittal values (i.e. 8.5 cm^2 vs 6.8 cm^2 for a 2.1 cm width). The lower pharynx values ('level 2') are also larger (i.e. 5.3 cm^2 vs 3.2 cm^2 for a 1.9 cm width). The larynx tube appears rather narrow. It may be worth noting that the laryngeal tube mainly controls the frequency of F_4, which was not corrected for the effect of the open glottis phase during a period.

It is not clear whether 'the relationship of lateral to cross-dimensions is affected by the phonetic nature of each vowel' and whether a tongue gesture factor, computed from the sagittal shape and related to the high/low and front/back features, is necessary to convert sagittal to area values, as suggested by Maeda (1971). While it is true that the formant fit is sometimes poor, the mean error is only 2.5 %. We have recently tested the 5 following vowels [i, e, oy, ø, œ], from a new X-ray film, using the coefficients given above. Mean percentage errors for the first to the fourth formants are 7.3, 9,7, 6.1, and 7.5 % respectively. These larger errors indicate that the set of coefficients

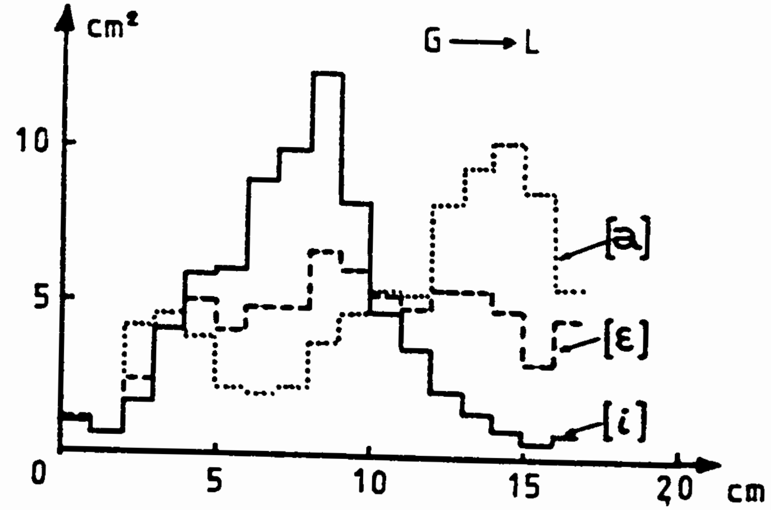

Figure 5. Area functions from the conversion of sagittal widths of x-ray data: vowels [i, ε, a].

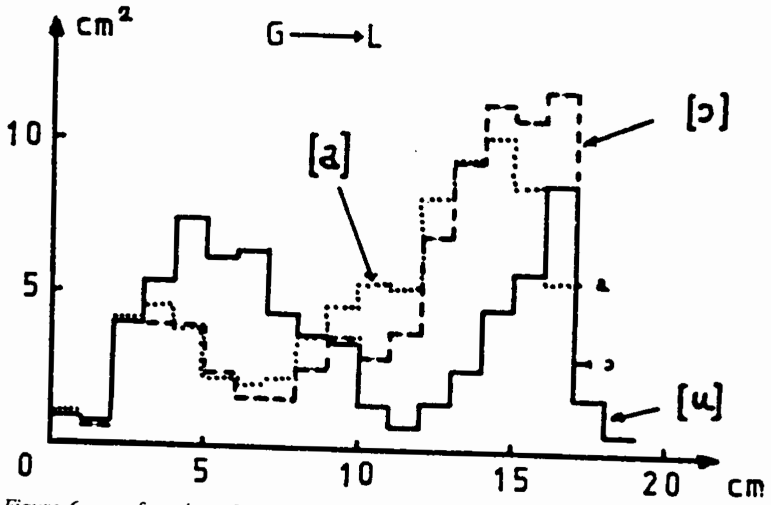

Figure 6. area functions from the conversion of sagittal widths of x-ray data: vowels [a, ɔ, u].

is not optimal. But the possibility exists that forcing area values to be a monotonically increasing function of the sagittal widths generates 'pseudo-area functions' with too strong a constriction in the larynx area. A multistage optimization procedure as developed by Charpentier (1982), combining table look-up for initial estimates and an optimization algorithm, might seem a better approach. But in this case, only one vowel can be dealt with at a time, as there are no constraints on the respective shapes of a set of vowels. Also, in order to have more acoustic variables than articulatory ones, one has to resort to measures of formant bandwidths or amplitudes, the relevance of which is doubtful. In our approach it can be checked that no sharp discontinuity exists for contiguous sections computed with different functions (i.e. at the junction between the mouth and upper pharynx functions) and that the area variation is physically compatible with the variation in sagittal width at most or all points along the vocal tract.

Further work, including simultaneous X-ray and acoustic impulse measurements as well as sagittal to area values conversion for a larger set of vowels, is definitely needed to better estimate area functions.

References

Charpentier, F. (1982). Application of an optimization technique to the inversion of an articulatory speech production model. *IEEE-ICASSP* 1982-1987.

Gauffin, J. and Sundberg, J. (1978). Pharyngeal constrictions. *Phonetica* **35**, 157-168.

Lefèvre, J.P., Tousignant, B. and Lecours, M. (1981). Utilisation de méthodes acoustiques pour l'évaluation des fonctions d'aires du conduit vocal. *XII^e Journées d'Etude du GALF*, Montréal, 26-41.

Lefèvre, J.P., Tousignant, B. and Lecours, M. (1983). Etude des configurations vocaliques des voyelles françaises à partir de mesures acoustiques. *Acustica* **52(4)**, 227-231.

Liljencrants, J. and Fant, G. (1975). Computer program for V.T. resonance frequency calculations. *STL-QPSR* **4** 1975, 15-20.

Maeda, S. (1971). Conversion of midsagittal dimensions to vocal-tract area function. *JASA* **51**, 88A.

Maeda, S. (1982). A digital simulation method of the vocal tract system. *Speech Communication* **1(3/4)**, 199-229.

Sondhi, M.M. (1979). Estimation of vocal tract areas: the need for acoustical measurements. *IEEE-ASSP* **27/3**, 268-276.

Sondhi, M.M. and Gopinath, B. (1971). Determination of vocal tract shape from impulse response at the lips. *JASA* **49**, 1867-1873.

Sondhi, M.M. and Resnick, J.R. (1983). The inverse problem for the vocal tract: numerical methods, acoustical experiments, and speech synthesis. *JASA* **73(3)**, 985-1002.

Tousignant, B., Lefèvre, J.P. and Lecours, M. (1979). Speech synthesis from vocal tract area function acoustical measurements. *IEEE-ICASSP*, 921-924.

Wood, S. (1982). X-ray and model studies of vowel articulation. *Lund University Working Papers n°* **23**.

Zerling, J.P. (1979). *Articulation et coarticulation dans les groupes occlusive-voyelle en français.* Thèse de 3ème cycle, Nancy.