THE MEASUREMENT OF THE ACOUSTIC TRANSFER FUNCTION AND THE AREA FUNCTION OF THE VOCAL TRACT : METHODS AND LIMITATIONS

B. Guérin

Institut de la Communication Parlée, Grenoble, France

ABSTRACT

In order to gain a greater insight into speech production phenomena, a better knowledge of both vocal tract acoustic transfer function and area function is needed. In these two kinds of investigation, many techniques can be used. We are going to analyse the different methods allowing us to obtain the correct results and their limits.

1. MEASUREMENTS OF VOCAL TRACT TRANSFER FUNCTION

Classical techniques for direct investigation of the acoustic characteristics of the vocal tract are based on the transcutaneous excitation of the tract near the glottis. The first measurements have been reported by Van den Berg [3]. Fant [10] later adopted a similar method. Well know results have been given by Fujimura & Lindquist [12] and this improved method is called sweep-tone measurement. In these different experiments, they used a pure tone signal swept in frequency as excitation. The source can be a small loudspeaker or a high-quality moving-coil-type electromagnetic transducer. In a session of data acquisition, a microphone picked out the sound at the mouth opening. The subject held the intended articulation as constant as possible during sweeping. A sweeping from 100 to 5 000 Hz took about 8,5 s. The major advantage of this method is the continuous frequency response curves that can be obtained. But three main disadvantages remain :

- The complex and particular experimental set up.
- A too long measurement time.
- No phonation is allowed during the process, but auditory feedback may be necessary for precise articulation configuration.

In order to resolve some of these problems, an improved method has been proposed by Castelli & Badin [5] [6] : the vocal tract was excited with white noise, which allows for auditory control. The output signal picked up by the microphone is directly sent to headphones : thus the subject can "hear" the configuration he is articulating. The signal is at the same time digitally recorded and further processed by averaging FFT spectra over a long period of time to produce good transfer functions. But the time measurement is still too big (around 10 s) and no phonation is possible.

A last improvement is proposed by Djeradi et al [9]. A technique developped earlier for acoustic room characterization has been adapted. The method is based on the impulse response of the vocal tract when it is excited throught the skin near the larynx. It is very difficult to produce a good impulse for excitation, so we use a pseudo-random sequence as input signal. The computation of the cross-correlation between the excitation and the output signal allows us phonation during the measurment process. The transfer function is obtained directly through the FFT of crosscorrelation. Finally, the measurement time is as small as about 100 ms, and the frequency resolution on the final result is of 10 Hz. As shown on figure 1, we can measure the transfer function with closed glottis condition and in phonation condition. With this method, we can consider measurement of the transfer function when the vocal tract is slowly moving, for example during vowel-vowel transition.

In conclusion, it is clear we can now measure the transfer function of the vocal tract with the following features :
- measurement time : about 100 ms
- accuracy in the frequency domain : about 10 Hz
- feasable measurement during phonation
- possibility to follow the variation of the transfer function during slow transitions

2. MEASUREMENTS OF THE AREA FUNCTION

The determination of articulatory parameters such as area functions or other representations of vocal tract shape is a long standing problem in speech research. This problem can be viewed accordingly in two ways : 1/ By direct measurement of the area using cineradiography, magnetic resonance imaging, ultrasonic scan, or X-ray microbeam. 2/ as an inverse problem from natural signal or output signal of external excitation. We are going to recall these different techniques and their limitations.

2.1. Direct measurement of vocal tract area

- X-ray photography an cineradiograph.

The first X-ray studies took place around 1925-1930. But the first well know were those from a Russian made by Fant in 1951. We obtained a mid-lateral saggital view of the vocal cavity. From the photography, we must first draw the outline of the vocal tract and after using a grid for example, we can measure the saggital distance of each cross section in order to define the centre line and the length of the vocal tract. Many errors occur during the processing of this kind of date :
- The boundary of the vocal tract is not always easy to define, in particular near the larynx.
- The choice of the method of outline to cross section magnitude conversion may give different values, especially on the total length of the vocal tract.

But a last operation must be done : the saggital distance → area cross - section conversion. Indeed the cross section shape of the vocal tract is very different according to position along the centre-line and even with the value of saggital distance. Usually, we use the Heinz & Stevens models where $A = \alpha d^{\beta}$. Computed tomography allowed us to obtain data of these cross-section shapes and a set of coefficients α and β were determined for several specific regions in the vocal tract [31]. Considering all the difficulties of this method, X-ray measurement, now improved with the cineradiographic records, gives very good information about the size and the shape of the vocal tract [17] [1] [4]. However, its use is restricted by the great deal of time and tedious work of processing involved.

Magnetic Resonance Imaging (MRI) [11] is a technique which allows the same kind of study as the X-ray method. Contrary to X-ray imaging, there are no severe problems owing to dosage limitations, but, for the moment, relatively long acquisition time make it impossible to investigate many speech sounds.

X-ray microbeam and ultrasonic scan often allow only the definition of a part of the vocal tract, usually the dorsal surface of the tongue [30] [26] [41].

2.2. Estimation of the area function from the speech signal

Considering that it is difficult to obtain the cross-sectional area function of the vocal tract from direct measurements, many researchers have tried to estimate it from acoustic data.

The first investigations were based on the inverse problem : determination of the vocal tract shape from transfer function or formant frequencies [32] [18] [20] [13] [2] [29]. But, as recalled by Mermelstein, Schroeder, Heinz and Sondhi [35], it is well known that the transfer function of a lossless vocal tract (much less that of a tract with loss) does not uniquely specify it's area function. An alternative is to try to resolve the non uniqueness by imposing "reasonable" constraints on the tongue and other articulators. But nothing allows us to judge if the result is the actual area function and not just any function which will produce the same sound.

Another approach was proposed, notably Wakita, Nakajima et al [44] [45] [46] [27], who estimated area functions from speech signals. The method is based on linear prediction analysis. If corrections on formant frequencies and bandwiths are made in order to compensate for the differences in boundary conditions and losses between the linear prediction model and actual speech production, reasonably results are obtained. Several problems remain in both of these approaches :

- the area function can be determined only to within an unknown factor,
- the vocal tract length is not directly available from the speech signal or the transfer function.

Wakita, Paige and Kirlin [28] [19] proposed algorithms to estimate this length and the various estimators give an accuracy of better 5 %. The approach which seems the most promising was proposed first by P. Mermelstein, and consists of considering the impedance function of the vocal tract measured at the lips. It has been shown that this characteristic allows the definition of a unique relation between the input impedance at the lips and the cross-section area function of the vocal tract. Sondhi has studied different aspects of the determination of area function [36] [37] [38] [39] [43] in depth. He shows that under plane wave approximation for a lossless vocal tract, as well as for tracts with certain types of distributed losses, we can reconstruct the area function with good concordance between calculated results and actual shape. Furthermore, J.P. Lefèvre, R. Descout, et al [8] [42] [22] used measurements at the lips of response to an impulse acoustic pressure wave for determining the vocal tract area function. This area function was obtained either by deconvolution, or by successive approximations of a modeled vocal tract, the search being made for the constrictions by decreasing order. In all cases, 20 to 40 measurements per second can be obtained.

In order to both improve the accuracy of the reconstructed area function and increase the frequency of measurement, P. Milenkovic [24] [25] has proposed a novel aspect of the procedure based on acoustic pulse reflection of an excitation at the lips. The results show that the period can be as short as 5 ms but only if articulatory is not too big.

These methods should be improved in different ways but two constraints remain : 1/ all of these need an impedance tube which is "connected" to the lips by a flexible coupling. There is bound to be some kind of unnaturalness in the articulatory movements. 2/ all the measurements must be made without phonation. This point could be neglected in the case of vocalic configurations, but it is very important for the acoustic characteristics of sounds like fricatives.

3. CONCLUSION

Different methods used for measuring the acoustic transfer function give accurate results. The determination of area functions is more difficult. Here different techniques are used and these give natural and verifiable results, using the driving point impulse response. Furthermore, much current research aims to obtain the area functions from the speech signal in order to achieve an efficient coding for transmission [21] [33] [34] and neural networks should also offer a new line of investigation [40].

Two ways should be explored (or continued) for increasing the quality of results : 1/ a better time model of the vocal tract, 2/ a mixed-method allowing the simultaneous measurement of transfer function and area function. Each of these studies have a specific application, firstly when we use a speech signal as input, and secondly when we use a synthetic external excitation.

REFERENCES

[1] AUTESSERRE, D., ROSSI, M., SARRAT, P., GIRAUD, G., VISQUIS, R., DEMANGE, R., CHEVROT, L. (1979), "Exploration radiologique de l'orpharynx, de l'hypopharynx et du larynx en phonation", *Proc. Séminaire Larynx et Parole (GRENOBLE)*, 45-74.

[2] BECKMANN, D.A., WOLD, D.C., MONTAGUE, J.C. (1983), "A nominuasive acoustic method using frequency perturbations and computer-generated vocal-tract shapes", *J.S.H.R., 26*, 304-314.

[3] VAN DEN BERG, S. (1955), "Transmission of the vocal cavities", *J.A.S.A., 27*, 161-168.

[4] BOTHOREL, A. (1984), "Apport de la radiocinématographie à la recherche phonétique. Phonétique instrumentale et linguistique", *Journées d'Etudes du 15 mai 1984*, 55-88.

[5] CASTELLI, E., BADIN, P. (1988), "Vocal tract transfer functions measurements with white noise excitation. Application to the naso-pharyngeal tract", *7th FASE Symposium (EDINBURGH)*, 415-422.

[6] CASTELLI, E., PERRIER, P., BADIN, P. (1989), "Acoustic considerations upon the low nasal formant based on nasopharyngeal tract transfer function measurements", *Proc. EUROSPEECH (PARIS), 2*, 412-415.

[7] CHARPENTIER, F. (1982), "Application of an optimization technique to the inversion of an articulatory speech production model", *Proc. ICASSP*, 1984-1987.

[8] DESCOUT, R., TOUSIGNANT, B., LECOURS, H. (1976), "Vocal tract area function measurements : two

time-domain methods", *Proc. ICASSP*, 75,78.

[9] DJERADI, A., GUERIN, B., BADIN, P., PERRIER, P. (1991), "Measurement of the acoustic transfer function of the vocal tract : a fast and accurate method", A paraître dans *Journal of Phonetics*.

[10] FANT, G. (1962), "Formant Bandwith Data", *STL Q.P.S.R., 1*, 1-2.

[11] FOLDVIK, A.K., HUSBY, O., KUAERNESS, J. (1988), "Magnetic resonnance imaging", *7th FASE Symposium (EDINBURGH)*, 423-428.

[12] FUJIMURA, O., LINDQVIST, J. (1970), "Sweep-tone measurements of vocal-tract characteristics", *J.A.S.A., Vol. 49, 2*, 541-558.

[13] FUJISAKI, H., OBATA, S., TAZAKI, R. (1971), "Estimation of vocal tract area function from poles of this transfer function", *Annual Report of the Engineering Research Institute, University of Tokyo, 30*, 81-88.

[14] GOPINATH, B., SONDHI, M.M. (1970), "Determination of the shape of the human vocal tract from acoustical measurements", *B.S.T.J., 49*, 1195-1214.

[15] GENIN, J. (1977), "Extraction de paramètres du conduit vocal à l'aide d'une méthode d'optimisation appliquée à un modèle mathématique", *Proc. Articulatory Modeling Symposium (GRENOBLE)*, 41-64.

[16] HEINZ, J. (1967), "Perturbation functions for the determination of vocal tract area functions from vocal-tract eigenvalues", *STL Q.P.S.R., 1*, 1-14.

[17] JOHANSSON, C., SUNDBERG, J., WILBRAND, H., YTTERBERGH, C. (1987), "From sagittal distance to area. A study of transverse, cross-sectional area in the pharynx by means of computed tomography", *Phonetica, 2*.

[18] KADOKAWA, Y., SUZUKI, J. (1969), "A simple calculation method for the vocal-tract configuration from the first three formant frequencies", *Electronics and Communications in Japan, Vol. 52-C, 9*, 123-120.

[19] KIRLIN, R.L. (1978), "A Posteriori Estimation of vocal tract length", *I.E.E.E. Trans. ASSP, Vol. 26, 6*,571-574.

[20] LADEFOGED, P., HARSHMAN, R., GOLDSTEIN, L., RICE, L. (1978), "Generating vocal tract shapes from formant frequencies", *J.A.S.A., Vol. 64, 4*, 1027-1035.

[21] LARAR, J.N., SHROETER, J., SONDHI, M.M. (1988), "Vector quantization of the articulatory space", I.E.E.E. *Trans. ASSP, Vol. 36, 12*, 1812-1818.

[22] LONCHAMP, F., ZERLING, J.P., LEFEVRE, J.P. (1983), "Estimating vocal tract area functions : A progress report", *Proc. Xth ICPhs (UTRECHT)*, 277-283.

[23] MERMELSTEIN, P. (1967), "Determination of the vocal-tract shape from measured formant frequencies", *J.A.S.A., Vol. 41, 5*, 1283-1294.

[24] MILENKOVIC, P. (1987), "Acoustic tube reconstruction from noncausal excitation", *I.E.E.E. Trans. ASSP, Vol. 35, 8*, 1089-1100.

[25] MILENKOVIC, P. (1984), "Vocal tract area funtions from two-point acoustic measurements with formant frequency constraints", *I.E.E.E. Trans. ASSP, Vol. 32, 4*, 1122-1135.

[26] MINIFIE, F.D., KELSEY, C.A., ZAGZEBSKI, J.A. (1971), "Ultrasonic scans of the dorsal surface of the tongue", *J.A.S.A., Vol. 49, 6*, 1857-1860.

[27] NAKAJIMA, T., OMURA, H., TANAKA, K., ISHIZAKI, S. (1974), "Estimation of vocal tract area functions by adaptative inverse filtering methods and identification of articulatory model", *Proc. Speech Communication Seminar (STOCKHOLM)*, 11-20.

[28] PAIGE, A. (1970), "Calculation of vocal tract length", *I.E.E.E. Trans. AU,*

Vol. 18, 3, 268-270.
[29] PAIGE, A. (1970), "Computation of vocal tract area functions", *I.E.E.E. Trans. AU, Vol. 18, 1*, 7-18.
[30] PERKELL, J.S., OKA, D.K. (1980), "Use an alternating magnetic field device to track midsagittal plane movements of multiple points inside the vocal tract", *99th Meeting of the A.S.A.*.
[31] PERRIER, P., BOË, L.J. (1987), "Passage de la coupe sagittale à la fonction d'aire", *16èmes J.E.P. (HAMMAMET)*, 128-131.
[32] SCHROEDER, M.R. (1967), "Determination of the geometry of the human vocal tract by acoustic measurements", *J.A.S.A., Vol. 41, 1*, 1002-1010.
[33] SCHROETER, J., LARAR, J.N., SONDHI, M.M. (1988), "Multi-frame approach for parameter estimation of a physiological model of speech production", *Proc. ICASSP*, 83-86.
[34] SCHROETER, J., SONDHI, M.M. (1989), "Dynamic Programming search of articulatory codebooks", *Proc. ICASSP*, 588-591.
[35] SONDHI, M.M., GODINATH, B. (1971), "Determination of vocal-tract shape from impulse response at the lips", *J.A.S.A., Vol. 49, 6*, 1867-1873.
[36] SONDHI, M.M. (1977), "Estimation of vocal-tract areas : the need for acoustical measurements", *Proc. Articulatory modeling symposium (GRENOBLE)*, 77-87.
[37] SONDHI, M.M. (1979), "Estimation of vocal-tract areas : the need for acoustical measurements", *I.E.E.E. Trans. ASSP, Vol. 27, 3*, 268-273.
[38] SONDHI, M.M. (1979), "Two acoustical inverse problems in speech and hearing", *Proc. Conf. Math. Methods Appl. Scattering Theory (WASHINGTION DC)*, 290-300.
[39] SONDHI, M.M., RESNICK, J.R. (1983), "The inverse problem for the vocal tract : numerical methods, acoustical experiment and speech synthesis", *J.A.S.A., Vol. 73, 3*, 985-1002.
[40] SOQUET, A., SAERENS, M., JOSPA, P. (1990), "Acoustic-articulatory inversion based on a neural controller of a vocal-tract model", *Proc. of the ESCA Workshop on Speech Synthesis (AUTRANS)*, 71-74.
[41] THOMPSON, M.A., ROBL, P.E. (1982), "X-ray microbeams for speech research", *Nuclear Instruments and Methods, 193*, 257-259.
[42] TOUSIGNANT, B., LEFEVRE, J.P., LECOURS, M. (1979), "Speech synthesis from vocal tract area function acoustical measurements", *Proc. ICASSP*, 921-924.
[43] VEMULA, N.R., ELLIOTT, D.L., ENGEBRETSON, A.M. (1982), "Estimation of vocal tract shape from input/output measurements", *Proc. ICASSP*, 927-931.
[44] WAKITA, H. (1983), "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveform", *I.E.E.E. Trans. AU, Vol. 21, 5*, 417-427.
[45] WAKITA, H., GRAY, A.H. (1975), "Numerical estimation of the lip impedance and vocal tract area functions", *I.E.E.E. Trans. ASSP, Vol. 23, 6*, 574-580.
[46] WAKITA, H. (1977), "Some considerations for the determination of area functions from acoustic date", *Proc. Articulatory Modeling Symposium (GRENOBLE)*, 89-96.