ELSEVIER

Medical Engineering & Physics 24 (2002) 467–471

Medical Engineering & Physics

www.elsevier.com/locate/medengphy

A method for measurement of the vocal tract impedance at the mouth

Malte Kob [a,*,1], Christiane Neuschaefer-Rube [b]

[a] *Institute of Technical Acoustics, Aachen University, Templergraben 55, D-52056 Aachen, Germany*
[b] *Phoniatric and Pedaudiologic Department, University Hospital, Aachen University, Pauwelsstraße 30, D-52074 Aachen, Germany*

Received 29 April 2002; accepted 24 May 2002

Abstract

In this contribution a method is presented for the measurement of vocal tract resonances. The technique uses a non-invasive acoustic excitation of the vocal tract and a fast and robust detection. The method is an alternative to the linear predictive coding (LPC) analysis for patients with voice and speech disorders. Sweep signals are emitted and recorded simultaneously from the small end of a tube placed in front of the mouth opening. The use of a pressure sensor and a velocity sensor provides a direct measurement of the vocal tract impedance at the mouth (VTMI). For selected sustained German vowels, and some consonants, a comparison of results from LPC analysis and VTMI measurements is given. The results indicate a good agreement in the frequency range from 500 to 5000 Hz. The feasibility of the VTMI method for diagnostic and therapeutic applications is subject to current research. © 2002 IPEM. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Acoustic impedance; VTMI; Vocal tract; Correlation technique; Velocity

1. Introduction

Phonemic speech information transmitted by acoustical signals is encoded by the time-dependent acoustic response of the vocal tract that is formed by different articulatory organs such as lips, tongue, velum and larynx. In order to define the phonemic positioning of these articulatory organs, X-Ray and magnet resonance imaging (MRI) techniques as well as ultrasonography [1,2] have been applied. However, some of those procedures are very costly. Alternative acoustical approaches are available to estimate the vocal tract transfer function (VTTF) [3] or even the vocal tract shape [4], although in an indirect manner.

The vocal tract impedance at the mouth (VTMI) presented in this paper is a characteristic descriptive measure of vocal tract acoustics like the vocal tract transfer function, c.f. [5]. Both measures can be used to derive the frequencies and bandwidths of the vocal tract resonances. Whereas a direct determination of the VTTF requires an in-situ measurement of the glottal sound pressure, the impedance method is non-invasive. Indeed the vocal tract impedance measured at the mouth has been used previously to characterise the resonant frequencies of the vocal tract [6], and to provide feedback on the vocal tract configuration as a training aid for the correct pronunciation of vowels [7,8].

2. Method

The VTMI method is based upon the set-up described in [6], which consists of a loudspeaker being attached to an impedance matching horn with an inserted high value acoustic resistor at its output end.

A schematic set-up of the new system for measurement of the vocal tract impedance at the mouth is shown in Fig. 1. A detailed description of the method is given in [9]. The authors of [6] used a high value acoustic resistor to ensure a close to ideal sound velocity source. Therefore for determination of the impedance only the pressure signal must be recorded at the end of the horn.

---

[*] Corresponding author.
E-mail address: mkob@ukaachen.de (M. Kob).
[1] Present address: Department of Phoniatrics, Pedaudiology and Communication Disorders, University Hospital, Pauwelsstraße 30, D-52074, Aachen, Germany.

PII: S1350-4533(02)00062-0

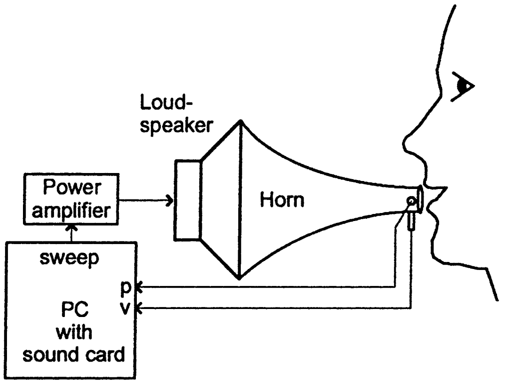


Fig. 1. VTMI measurement set-up.

This is done with a miniature microphone positioned close to the vestibulum oris (vestibule of the mouth) of the speaker under test.

Preliminary experiences with a low output impedance using a similar set-up as reported in [6] have shown that the velocity at the vestibulum oris was significantly altered by the load of the vocal tract, although horn and resistor should prevent such variations. Increasing the value of the acoustic resistor reduced the effect but also the output power and, consequently, the signal-to-noise ratio. The improvement of the new setting is to measure the velocity directly by a miniature sensor (Microflown, c.f. [10]) close to the microphone (Sennheiser KE 4 211-2), as depicted in Fig. 1. This has two advantages: 1) reduction of the power required at the loudspeaker and 2) miniaturization of the measurement set-up.

For measurement of the impedance, a technique using a swept sine as excitation signal is applied [11]. The signal is processed identically for both sensors following the signal flow given in Fig. 2.

FFT represents the fast Fourier transform and IFFT denotes the inverse fast Fourier transform. The idea of using the reference measurement is to compensate deviations of the measurement system from the ideal frequency transfer function. After division of the measured spectrum by the reference spectrum, the IFFT yields a time signal that equals the impulse responses of sound pressure $h_p(t)$ and velocity $h_v(t)$ at the mouth.

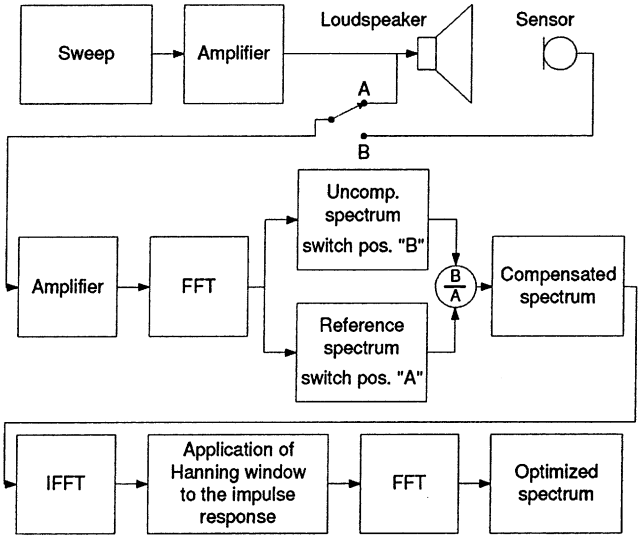


Fig. 2. Signal flow of impedance measurements.

The acoustic point impedance $Z(f)$ is defined as a quotient of pressure and velocity spectra

$$Z(f) = \frac{\mathrm{FFT}(h_p(t))}{\mathrm{FFT}(h_v(t))}. \tag{1}$$

An advantage of the sweep technique is that harmonic distortions can be cancelled out by windowing the impulse response (here: Hanning window, 0..20 ms). Thus, minor changes of the mouth position or the vocal tract configuration during measurement do not reduce the overall signal-to-noise ratio since the distortion only affects the frequency range corresponding to the instant when the configuration is changed. In the following, the impedances are displayed as magnitude $\left|\frac{Z}{Z_0}\right|$ of the ratio of the impedance Z of the acoustic system under test to the unloaded, free-field impedance Z_0. The upper frequency range of the velocity sensor used allows reading of resonance frequencies, relative amplitudes and quality of the resonances up to approx. 5 kHz. Using a PC (800 MHz), measurements with a frequency spacing of 2.7 Hz at a sampling rate of 44.1 kHz and a repetition rate below 1 Hz are possible.

The apparatus has been applied to the measurement of German long vowels and selected consonants (see Table 1). In order to control the steadiness of phonemic positioning of the articulatory organs during the mutely executed VTMI-measurements, ultrasonography of the tongue was carried out simultaneously with the other measurements. A 5MHz convex probe attached to the Hitachi EUB-405 sonograph was placed in the submental region imaging the air-soft part contour of the tongue dorsum. Details of this method were published previously [1,2]. All tongue positions of the examined phonemes were visualized in the mediosagittal plane.

3. Measurements and results

The resonance evaluation was improved by using the velocity sensor in addition to the pressure sensor, as presented in Fig. 3. The figure depicts the result of a measurement on an aluminium model [9] with an equivalent area function similar to the vocal tract configuration [Λ].

The thin line indicates the absolute value of the spectrum of the sound pressure measured in a position equivalent to the vestibulum oris, the dotted line shows the

Table 1
Mean values and standard deviations (in brackets) of formant frequencies of the vowels (F2 & F3/F4*) resp. poles of the consonants, obtained by four independent measurements using LPC and VTMI methods

Phoneme	Expression	LPC [Hz]	VTMI [Hz]	LPC [Hz]	VTMI [Hz]
Vowels		2nd Formant		3rd or 4th (*) Formant	
[aː]	**A**ber	781 (60)	1110 (15)	3176 (238)	3515 (85)
[ɛː]	**Ä**hre	1680 (86)	1520 (20)	3335 (157)*	3450 (20)*
[eː]	b**e**ten	2062 (51)	2450 (50)	3251 (42)	3250 (50)
[iː]	M**ie**te	2261 (118)	2600 (50)	3187 (65)	3275 (25)
[yː]	H**ü**te	1901 (67)	2225 (25)	3036 (126)*	3015 (35)*
Consonants		1st Pole		2nd Pole	
[ʃ]	Sa**sch**a	2266 (67)	2125 (75)	4816 (134)	4750 (50)
[x]	a**ch**	1076 (72)	1045 (25)	3742 (38)	3450 (250)
[j]	ja**j**a	3300 (60)	3300 (100)	4482 (235)	4150 (50)

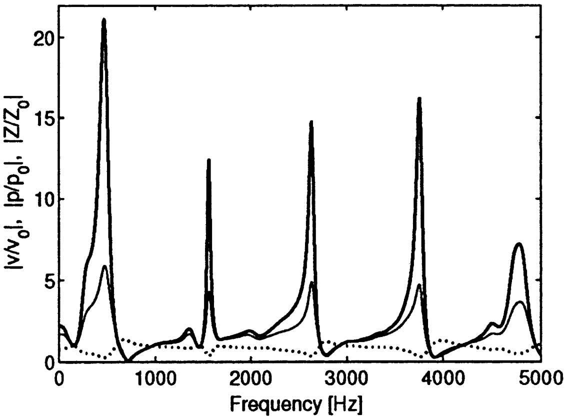


Fig. 3. Normalized velocity, sound pressure and impedance measured on a model.

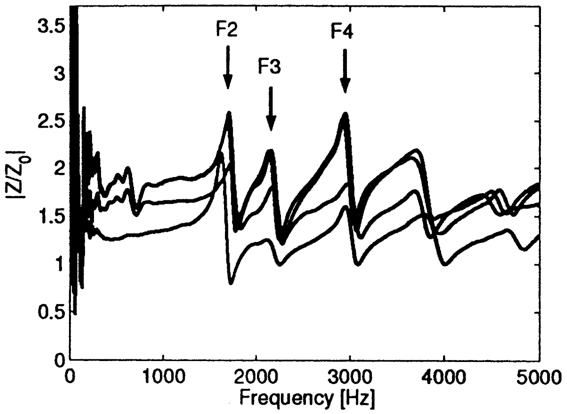


Fig. 4. Plot of four measurements of the VTMI ratio, vowel [yː].

absolute value of the velocity measured at the same location and the thick solid line represents the absolute value of the acoustic point impedance. The resonance frequencies can be identified as local maxima of the impedance curve and the sound pressure curve. It can be observed that the impedance curve yields a significantly higher peak amplitude than the pressure curve. This effect corresponds to the finding that the velocity at the mouth is significantly reduced at the resonance frequencies. The reproducibility of measurements on a human subject is shown in Fig. 4.

Four subsequent measurements have been carried out while the subject (male, age 33) articulated the vowel [yː]. The arrows indicate the measured resonance frequencies that identify the formants of the vocal tract. The resonances vary only a little between the measurements but some curves are shifted vertically with respect to others, indicating a different distance between measurement head and mouth opening.

Two other kinds of measurements were performed with the same subject. In the first experiment, the pronounced phoneme was recorded four times via a microphone close to the subject's mouth. The recording of the phonemes was evaluated by a linear predictive coding (LPC) analysis using the COLEA[2] package for the program language MATLAB. In the second measurement, the VTMI was measured four times. The results of the measurements are summarized in Table 1. In both measurements the subject was asked to hold the pronunciation of the expressions on the phonemes indicated in bold letters.

By sonographic observation it was verified that the articulatory configuration did not change between the measurements. However, the invariability of the vocal tract configuration could only be verified in the imaged plane. As a consequence of the small size of the loudspeaker used in the measurement set-up, only a determination of the resonances above 500 Hz could be achieved using the VTMI method. However, the use of a larger loudspeaker should allow a determination of the first formant [7,9].

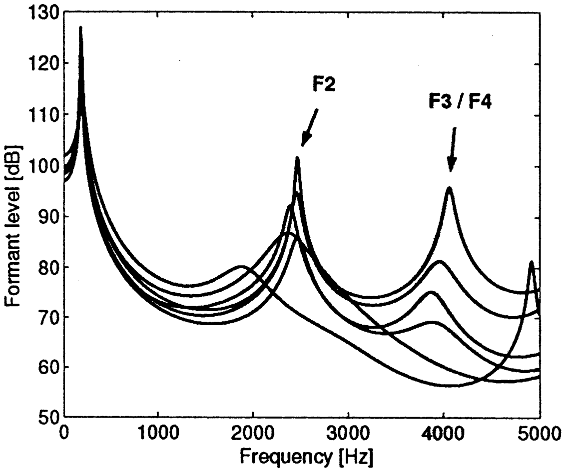
Another example of a comparison of the two methods

---

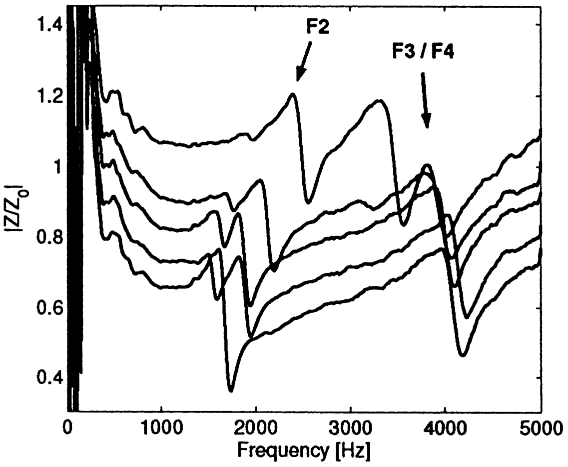
[2] Available on the internet at
http://www.utdallas.edu/~loizou/speech/colea.htm

is given in Fig. 5. The transition of formants of the vowel [iː] pronounced in several tongue positions changing successively from ventral (front position) to a more dorsal (back) position are presented as LPC analysis (a) and VTMI measurement (b). The upper curves in each plot represent the spectra of the vowel [iː] (ventral tongue position) whereas the lower curves point out a shift of F2 towards lower frequencies caused by a rather dorsal position of the tongue. The upper formants F3 and F4 are difficult to track using LPC analysis but remain visible using the VTMI method. The arrows indicate the peak shifts of the subsequent curves.

The curves from the LPC analysis of the voice signal represent a mathematical model that fits the measured data. The formants can be identified as local maxima at the resonance frequencies. The interpretation of the impedance curves is more difficult since they represent "raw" data. Due to the boundary condition at the mouth the sound pressure can be very small while the velocity is rather big, resulting in zeros in the impedance curve. Consequently, the LPC curves and the VTMI curves are substantially different.



(a) LPC analysis (order 8, 50 ms window)



(b) VTMI analysis, curves shifted by 0.1

Fig. 5. LPC (a) and VTMI (b) curves for a sequence of [iː] with tongue moved.

The values indicate a satisfactory reproducibility for each method with exception of a few phonemes. A comparison of LPC and VTMI analysis should result in similar, though not necessarily identical, values for the formant or pole frequencies. However, in some cases significant differences have been observed.

4. Conclusions

We have demonstrated that the VTMI method is able to visualize functional vocal tract characteristics during articulation. The method presented yields fast and reliable impedance measurements in the frequency range from ~500 Hz to 5 kHz, even with a single sample of 166 ms duration. However, results from LPC analysis and VTMI measurements seem to differ significantly for some phonemes. This may be caused by undetected tongue movements not visualized in the selected sonographic imaging plane, or by inherent differences between vocal tract resonances measured with or without phonation. The differences between VTMI measurements with glottis open, glottis closed and during phonation have been found not to be neglectable [9]. Further, a change of configuration between recording and VTMI measurement cannot be excluded with certainty. The problem should be solved by using an optimized set-up allowing an on-line voice recording and VTMI measurement in very fast sequence or even simultaneously, as described in [6].

Further studies are intended to work out different clinical applications of the method. Our preliminary results shown for the vowel [iː] indicate that the method is able to differentiate among various kinds of vocal tract configurations. With some additional knowledge about the configuration of vocal tract and glottis it should be possible to derive the VTTF and even the vocal tract area function from the VTMI measurement. Additional studies are intended to examine the impedance characteristics of the vocal tract in groups of patients suffering from hyperfunctional dysphonia compared to groups of healthy adults. The voice production of these patients is often characterized by an impaired vocal tract function caused by the pathological use of the false vocal chords or by a functional backward dislocation of the tongue. In addition, examinations in healthy children are proposed to be useful to obtain clinical data for a prospective application of the method in the articulatory training of deaf children. In this context, it is of great advantage that this method is able to visualize the vocal tract function even in mutism [7,8,9].

Acknowledgements

The authors would like to thank Bernd J. Kröger for his helpful comments on the manuscript and the results. The reviewers are gratefully acknowledged for their critical commments.

References

[1] Sonies BC, Shawker TH, Hall TE, Gerber LH, Leighton SB. Ultrasonic visualization of tongue motion during speech. J. Acoust. Soc. Am. 1981;70:683–6.

[2] Wein B, Böckler R, Huber W, Klajman S, Willmes K. Computer-sonographische Darstellung von Zungenformen bei der Bildung der Langen Vokale des Deutschen (in German). Ultraschall Med 1990;11:100–3.

[3] Fujimura O, Lindquist J. Sweep-tone measurements of vocal-tract characteristics. J Acoust Soc Am 1971;49:541–58.

[4] Sondhi MM, Gopinath B. Determination of vocal-tract shape from impulse response at the lips. J Acoust Soc Am 1971;49:1867–73.

[5] Pham Thi Ngoc Y., Charactérisation acoustique du conduit vocal: Fonctions de transfert acoustiques et sources de bruit. PhD Thesis, Inst Nat Polytech Grenoble, 1995 (in French).

[6] Epps J, Smith JR, Wolfe J. A novel instrument to measure acoustic resonances of the vocal tract during phonation. Meas Sci Technol 1997;8:1112–21.

[7] Dowd A, Smith J, Wolfe J. Real time, non-invasive measurement of vocal tract resonances: Application to speech training. Acoustics Australia 1996;24(2):53–60.

[8] Dowd A, Smith J, Wolfe J. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. Language and Speech 1997;41(1):1–20.

[9] Kob, M., Physical modeling of the singing voice. PhD Thesis, Aachen University. Berlin: Logos-Verlag, 2002.

[10] de Bree, H.-E., (2002) An overview of microflown technologies. Submitted to acta acustica—ACUSTICA, further information is available at URL http://www.microflown.com.

[11] Müller S, Massarani P. Transfer function measurement with sweeps. J Audio Eng Soc 2001;49(6):443–71.