Available online at www.sciencedirect.com
ScienceDirect
Computers and Electrical Engineering 34 (2008) 547–556
Computers and Electrical Engineering
www.elsevier.com/locate/compeleceng

Technical Communication

Modeling of vocal tracts based on polynomials

M.R. Karami Mollaei *, Malihah Hassani

Faculty of Electrical and Computer Engineering, Noushirvani Institute of Technology, University of Mazandaran, P.O. Box 484, Babol, Iran

Accepted 17 January 2007
Available online 26 March 2008

Abstract

A great problem in speech processing is to represent the shape and characteristics of the vocal tracts. This task is normally done by using an acoustics tube model, based on the calculation of the area function would be performed. The area function can be found from sagittal function determined by sagittal cub. In this paper, an optimization based method is presented to model the vocal tracts and their relationship with the fundamental frequencies of vowels and consonants. We present two models which are simple polynomials that need the minimum number of parameters. We will show that these models have good performance in experiments.

Keywords: Vocal tract; Acoustic; Area function; Sagittal function

1. Introduction

The early studies on the mechanisms of the sound production of words were based on modeling of the vocal tracts by an acoustic tube incited by one or several sources. This modeling that requires extracting the articulator parameters is commonly used to characterize the vocal tracts. The articulator parameters can be extracted from acoustic signals.

Unfortunately, the articulatory – acoustic relation is a many-to-one function and finding the inversion of such a function, is an ill-posed inverse problem. [7,3,14]. This means that, it is not certain that the existing solution is unique, and continuously dependent on the initial condition. Moreover, finding the articulator parameters from the speech, is a very challenging problem with many perspectives such as articulatory gesture, learning by a speech robot [17], speech perception, and speech low bit – rate coding [5,16].

A common technique to resolve most of ill-posed problems is to introduce a set of appropriate constraints. Among all the possible solutions, we obtain the solution that agrees with the constraint set and settle the present problem. Area functions are widely used to represent the shape of the vocal tracts. Two different approaches for deriving area functions are:

* Corresponding author.
E-mail addresses: mkarami@nit.ac.ir (M.R.K. Mollaei), Hassani_ma2002@yahoo.com (M. Hassani).

0045-7906/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compeleceng.2007.01.004

(1) Direct methods involving geometrical measurements of the vocal tracts.
(2) Indirect methods based on acoustic inversion.

Among the direct measurement methods, only "Magnetic Resonance Imaging" can provide a complete three-dimensional picture of the vocal tracts [6], X-ray tomography of the entire vocal tracts which is used in this method, being considered too potentially hazardous for any practical use.

The other direct methods provide partial information. X-ray radiography which is widely spread, delivers midsagittal profiles of the vocal tracts, [7,4,22]. Also we can be associated with optical lip recordings [10,4] computer-aided X-ray tomography can also provide very useful data [5,29,30,22,24].

Palatography or electropalatography furnishes essentially the contours of tongue contact with the hard palate [15], and is of limited interest because of the reconstruction of the vocal tracts shape, unless associated with ultrasound imaging techniques [26]. Ultrasound techniques provide information on the tongue surface shape [27], but only in restricted zones of the tongue. The realization of casts of the vocal tracts is a very informative method, limited by its invasive character [31].

Finally, optical measurements of the lips provide a few important parameters, such as intra labial area or height, width and depth of the labial horn [1,19,13].

Indirect methods are based on determining the area function from acoustic data, either from the speech signal or from the acoustic response of the vocal tract [23,25,28].

The difficulty of the measuring the area function directly, has often been evoked in the literature [8,11,9]. On the other hand, indirect methods can not guarantee to yield the true area function [21].

Several researchers focused on indirect methods involving the extraction of the area function from midsagittal profiles [7,20,17,6,4,29,30].

Most of reported experiments on indirect methods (Except [7]) have not exhibited good accuracy regardless of the fact that they worked only on vowels. This means the formants computed using the area functions are far from those measured by the midsagittal profile [32,12].

Therefore a combined approach has been adopted: a set of midsagittal profiles and corresponding harmonics has been measured, and a model based on these data has been developed to convert midsagittals to area functions.

With the sagittal functions, we can represent the space of entrance strictly bound to an experimentally measurable sagittal profile by radiography [30,18]. To perform a model to convert midsagittal function to area function and to ensure the coherence of the three sets of data (midsagittal functions, area functions and formants frequencies), on the whole set of configurations, an appropriate set of constraints is needed.

In this paper, our contribution is to combine both approaches, and to derive realistic area functions coherent with midsagittal profiles and its signal producing, through acoustic models. Our model is based on the strong constraint that can work for sound as different as vowels and consonants, and is coherent with both midsagittal and acoustic levels. Moreover, this method guarantees the reliability of the area function determination.

2. The passage from the sagittal cup to the area function

The Characterizing, a cylindrical pipe from its acoustic signals, has a great analogy to obtain the transfer function, using input-output electrical signals [18].

The main point is that the sagittal cup is two dimensions and the area function is three dimensions [17]. Now there is an important question, whether it is possible to reach from sagittal cup of the vocal tracts to its area function or not.

Numerous works in radiography and in radio cinematography have presented models of the sagittal cup of the vocal tracts [17,30].

The early works associating area function and transfer function of the vocal tracts go back to year 1941, by using moldings, predicting and feigning the Japanese vowels [7].

Currently, the measurements of area function for the continuous word are practically impossible so it would be necessary to arrange cups of area every 5 or 10 ms to aspire a good precision. Therefore indirect measurements were developed as follow [2,32]: