Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 33 (2015) 47–66
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Two dimensional featured one dimensional digital waveguide model for the vocal tract☆

Tahir Mushtaq Qureshi*, Khalid Saifullah Syed

Center for Advanced Studies in Pure and Applied Mathematics (CASPAM), Bahauddin Zakariya University, Bosan Road, Multan 60800, Pakistan

Received 10 October 2013; received in revised form 18 July 2014; accepted 15 December 2014
Available online 25 December 2014

Abstract

A vocal tract model based on a digital waveguide is presented in which the vocal tract has been decomposed into a number of convergent and divergent ducts. The divergent duct is modeled by a 2D-featured 1D digital waveguide and the convergent duct by a one dimensional waveguide. The modeling of the divergent duct is based on splitting the volume velocity into axial and radial components. The combination of separate modeling of the divergent and convergent ducts forms the foundation of the current approach. The advantage of this approach is the ability to get a transfer function in zero-pole form that eliminates the need to perform numerical calculations on a discrete 2D mesh. In this way the present model named as a 2D-featured 1D digital waveguide model has been found to be more efficient than the standard 2D waveguide model and in very good comparison with it in the formant frequency patterns of the vowels /a/, /e/, /i/, /o/ and /u/. The model has two control parameters, the wall and glottal reflection coefficients that can be effectively employed for bandwidth tuning. The model also shows its ability to generate smooth dynamic changes in the vocal tract during the transition of vowels.

PACS: 43.70.Bk; 43.70.Aj; 43.70.Ep

Keywords: Waveguide; Vocal tract; Fractional delay; Transfer function

1. Introduction

Human speech production system consists of three main components like lungs, vocal folds and vocal tract. The coordination of these three components results into voiced sound, unvoiced sound or combination of these two. For voiced sound production like that of vowel, the air is pushed out from the lungs into the larynx. In the larynx, there are two identical vocal folds which are initially closed. The closure of the vocal folds causes a sub-glottal pressure. When this pressure rises above the resistance of the vocal folds, the vocal folds open themselves and air is passed through it. As the pressure decreases with the release of airflow, the vocal folds then close themselves quickly. The quasi-periodic opening and closing of the vocal folds continues due to constant supply of the air pressure from the

☆ This paper has been recommended for acceptance by P. Jackson.

* Corresponding author. Tel.: +923146110616.
E-mail address: tahmush@hotmail.com (T.M. Qureshi).

http://dx.doi.org/10.1016/j.csl.2014.12.004
0885-2308/

lungs. Thus the vibration of the vocal folds forms a train of periodic pulses that acts as an excitation signal for the vocal tract. A non-uniform acoustic tube which extends from the glottis to the lips is called a vocal tract. The position of the vocal articulators like larynx, velum, jaw, tongue, and lips, forms a particular shape of the vocal tract. The shape of the vocal tract modifies spectral characteristics of the quasi-periodic air flow passing through it, which leads to the generation of voiced speech. In this way different shapes of the vocal tract generate different voiced speeches.

Several approaches have been employed to model the voiced speech system on the basis of physical models such as cylindrical segments (Kelly and Lochbaum, 1962; Mullen et al., 2003) and conical segments (Välimäki and Karjalainen, 1994; Strube, 2003; Makarov, 2009) for the vocal tract modeling. In cylindrical approach, each tube segment of the vocal tract is modeled by the forward- and backward-traveling wave components of the solution of the wave equation (Morse, 1981; Smith, 1998) known as one-dimensional waveguide model. It was firstly used in Kelly–Lochbaum model of the human vocal tract for speech synthesis (Kelly and Lochbaum, 1962). However, the digital waveguide modeling (DWM), which is an extension of a one-dimensional waveguide, is recently being used in the modeling of the vocal tract (Van Duyne and Smith, 1993a,b; Cooper et al., 2006; Mullen et al., 2006, 2007; Speed et al., 2013). Digital waveguides are very popular for realistic and high quality sound generation in real time, and are successfully employed in physical modeling of sound synthesis.

The greatest advantage of a 1-D digital waveguide model is that it has complete solution to the wave equation which is also computationally efficient for sound synthesis applications. Moving to higher dimensions leads to a number of limitations imposed on DWM models for an optimal solution to all sound synthesis systems. The most important one is the dispersion error, where the velocity of a propagating wave depends upon both its frequency and direction of traveling, leading to wave propagation errors and mistuning of the expected resonant modes. The dispersion error is highly dependent upon mesh topology and has been investigated in (Van Duyne and Smith, 1996; Fontana and Rocchesso, 2001; Campos and Howard, 2005). Another limitation is the restriction on sampling frequency. High sampling rates require high mesh density which corresponds to high computational cost.

A 1D waveguide model is computationally efficient while the standard 2D and 3D waveguide models have better accuracy but heavy computational cost (Murphy and Howard, 2000; Campos and Howard, 2000; Beeson and Murphy, 2004; Murphy et al., 2007). In the present work we propose an efficient two-dimensional waveguide model of the vocal tract that has comparable formant frequencies with the standard 2D waveguide but has efficiency comparable to that of a 1D waveguide model. In the present model we approximate only the divergent part of the vocal tract by divergent ducts and consider two-dimensional volume velocity in it while in the convergent duct that represents convergent part of the vocal tract, we employ conventional one-dimensional approximation of the volume velocity. In this way the accuracy of the current model can never be better than the standard 2D waveguide model which considers two-dimensional volume velocity in the whole of the vocal tract. Therefore, we make it as a reference model for the comparison.

The present results of the formant frequencies from the numerical simulation using area functions for specific vowels (Juszkiewicz, 2014) exhibit good comparison with the standard 2D waveguide model. The computational cost of the standard 2D waveguide is very high while the current approach is much more efficient.

The present section is followed by five more sections. In Section 2, we describe our proposed vocal tract model. In this section, we also develop its mathematical formulation. Section 3 describes how to find a transfer function of the vocal tract. Section 4 is reserved for the numerical simulation of the model. Section 5 is dedicated for the results and discussion and Section 6 is for the conclusions.

2. Vocal tract model

We derive a new model of vocal tract with a new transfer function relating it to pole-zero type linear prediction developed on the basis of the procedure given in (Kang and Lee, 1988). Current approach is to propose an efficient two-dimensional waveguide that has formant frequencies comparable with those of the standard 2D waveguide. We consider the vocal tract consisting of concatenated cylindrical acoustic tubes of same lengths but different cross-sectional areas. We define a convergent duct by the concatenation of two cylinders, where a cylinder with larger radius is followed by the one with the smaller radius. The connection of two cylinders in which a narrow cylinder is followed by a wider cylinder in the direction of flow is called a divergent duct. A serial combination of these two types of ducts constitutes the vocal tract. For example, in Fig. 1, the concatenation of the cylinders l_1 and l_2 forms a divergent duct while that of