THE DYNAMICALLY VARYING DIGITAL WAVEGUIDE MESH

PACS: 43.75.Zz

Murphy, Damian¹; Shelley, Simon²; Ternström, Sten³; Howard, David⁴
¹Intelligent Systems – Audio Lab; Department of Electronics, University of York, Heslington, York, YO10 5DD, UK; dtm3@ohm.york.ac.uk
²Intelligent Systems – Audio Lab; Department of Electronics, University of York, Heslington, York, YO10 5DD, UK; sbs107@ohm.york.ac.uk
³KTH CSC – TMH: Department of Speech, Music and Hearing, Kungliga Tekniska Högskolan, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden; stern@csc.kth.se
⁴Intelligent Systems – Audio Lab; Department of Electronics, University of York, Heslington, York, YO10 5DD, UK; dh@ohm.york.ac.uk

ABSTRACT
The digital waveguide mesh (DWM) is a multi-dimensional numerical simulation technique used to model vibrating objects capable of supporting acoustic wave propagation, with the result being sound output for excitation by a given stimulus. To date most DWM based simulations result in the static system impulse response for given initial and boundary value conditions. This method is often applied to room acoustics modelling problems, where the offline generation of impulse responses for computationally large or complex systems might be rendered in real-time using convolution based reverberation. More recently, work has explored how the DWM might be extended to allow dynamic variation and the possibility for real-time interactive sound synthesis. This paper introduces the basic DWM model and how it might be extended to include dynamic changes and user interaction as part of the simulation. Example applications that make use of this new dynamic DWM are explored including the synthesis of simple sound objects and the more complex problem of articulatory speech and singing synthesis based on a multi-dimensional simulation of the vocal tract.

INTRODUCTION
The digital waveguide mesh (DWM) [1] is a multi-dimensional numerical simulation technique based on the definition of a regular spatial sampling grid for a particular problem domain. The spatial sampling points are generally termed scattering junctions, as an incoming signal will be distributed (scattered) according to the number of connections and their relative impedances. Scattering junctions are connected via bi-directional waveguide elements facilitating lossless signal transmission between sampling points. This approach, and the terminology used, are derived from digital waveguide sound synthesis [2], a time and space discretization of the d’Alembert solution to the 1-D wave equation. This method was first used in the Kelly-Lochbaum model of the human vocal tract for speech synthesis [3] and is now a highly popular and successful technique in the field of computer music. This is due to the realistic, high quality sounds that can be generated with the potential for real-time implementation, therefore facilitating effective user interaction. Extending this technique to the multi-dimensional DWM case generally results in a non-realtime implementation where the output signal is the response of the modelled system to an appropriate stimulus. This offline generated impulse response is then convolved with an arbitrary input for audio processing and sound transformation applications. Whereas digital waveguides are efficiently implemented using bi-directional delay lines and applicable to many systems of interest to the computer musician, the DWM attempts to simulate the actual propagating medium of a multi-dimensional system such that the model is defined to be geometrically comparable to the system being studied. However most such DWM implementations are fixed according to model geometry, initial and boundary value conditions. Once excitation(s) have been applied, there is no interaction with the simulation, in part to ensure stability, but mainly due to the offline nature of the simulation. Therefore most DWM work to date has focused on application areas where an impulse response can be obtained from the model for appropriate post-processing, hence the interest in this method for room acoustics simulation e.g. [4].
This paper introduces a new formulation of the DWM that facilitates real-time dynamic variation and therefore offers the potential for user interaction. This new implementation has been previously investigated in the context of articulatory speech synthesis based on a simulation of the vocal tract as presented in [4] [5]. The next two sections of this paper briefly introduce the DWM, and how it might be adapted for a dynamically varying simulation. Results are produced using a time varying vocal tract model for the generation of simple natural sounding speech, sound synthesis using a time-varying membrane, and improved articulatory speech synthesis. The paper concludes by highlighting areas for future work.

THE DIGITAL WAVEGUIDE MESH

DWM structures are constructed using bi-directional delay lines and scattering junctions which act as spatial and temporal sampling points within the problem domain. Note that for a full derivation of what follows the reader is referred to [1] [2] [4]. The sound pressure in a waveguide element is represented by $p_i$, the volume velocity by $v_i$ and the impedance of the waveguide by $Z_i$ where $p_i/v_i = Z_i$. The admittance $Y_i$ is the inverse of $Z_i$, such that $Y_i = 1/Z_i$. The input to a waveguide is termed $p_i$ and the output $p_i$. The signal $p_{J,i}$ therefore represents the incoming signal to junction $J$ along the waveguide from the opposite junction $i$ and $p_{J,i}$ represents the outgoing signal from junction $J$ along the waveguide to the opposite junction $i$.

![Figure 1.- (a) Functional block diagram of a scattering junction $J$ with $N$ neighbours; (b) 4-port rectilinear DWM topology; (c) 6-port triangular DWM topology.](image)

Fig. 1(a) shows the functional block diagram of a scattering junction $J$ with $N$ neighbours with each connected unit waveguide element having an associated admittance $Y_i$. Fig. 1(b) shows how delay line elements might be connected together via 4-port scattering junctions to form a rectilinear spatial and temporal sampling grid. An alternative mesh topology based on 6-port scattering junctions is shown in Fig. 1(c). The 1-D waveguide can be implemented using two bi-directional (unit) delay lines as shown in the waveguide element of Fig. 1(a), hence the sound pressure of a propagating wave signal can be defined as the sum of these travelling waves or alternatively the input and output of this waveguide element:

$$p_j = p_{j,i}^+ + p_{j,i}^-$$  \hspace{1cm} (Eq. 1)

For a lossless junction $J$ the sum of the input velocities is equal to the sum of the output velocities, and the sound pressures in all crossing waveguides are equal, hence the sound pressure $p_j$ at junction $J$ for $N$ connected waveguides can be expressed as:

$$p_j = \frac{2 \sum_{i=1}^{N} Y_i \cdot p_{j,i}^+}{\sum_{i=1}^{N} Y_i}$$  \hspace{1cm} (Eq. 2)

Note that Eq. 2 can also be derived directly from Fig. 1(a). As the waveguides are equivalent to bi-directional unit-delay lines, the input to scattering junction $J$ is equal to the output from neighbouring junction $i$ into the connecting waveguide at the previous time step. Expressing this relationship in the z-domain gives:

$$P_{j,i} = z^{-1} \cdot P_{j,i}^-$$  \hspace{1cm} (Eq. 3)

Equations (1), (2) and (3), are termed the scattering equations for the mesh structure. Other finite-difference type formulations in terms of junction pressures only (K-variable or Kirchhoff formulations) are also possible. This leads to hybrid DWMs consisting of mixed wave-based and Kirchhoff-based scattering junctions, with such K-variable or hybrid approaches leading to more efficient simulations in terms of both memory use and computation time [6] [7].
One of the significant advantages of the 1-D digital waveguide that originally made it a realistic proposition for applications in sound synthesis is the computational efficiency of the approach when compared with a brute force numerical solution to the system wave equation. This is further improved through the ability to commute losses to specific lumped points in the system, significantly reducing the number of calculations required per time-step iteration. Unfortunately the elegance of this approach is lost when moving to higher dimensions. With a DWM based system, acoustic wave propagation is determined by signal interaction at the scattering junctions and hence a calculation must take place at every junction for every time-step. Reducing the number of scattering junctions reduces the sample rate of the DWM and hence the effective bandwidth of the system. The advantage gained with the DWM approach however is in the structural immediacy of the simulation, allowing objects to be defined based only on physical and geometrical definitions, and the ability to observe and interact with the system at physically relevant or meaningful points.

Dispersion error, where the velocity of a propagating wave is dependent upon both its frequency and direction of travel, must also be considered. The degree of dispersion error is highly dependent upon mesh topology and has been investigated in much of the related literature (see [4] for a comprehensive list). Frequency warping techniques can be used to correct these errors although with a sufficiently high sample rate ensuring accuracy at low frequencies where resonant frequencies are sparsely distributed, it is suggested that the perceptual effects of dispersion become less critical, although this has not been tested thoroughly as yet.

Finally note that the mesh sample rate is determined from the associated Courant stability condition such that \( f_{\text{update}}=(cD^{0.5})/d \), where \( f_{\text{update}} \) is the sample rate update frequency for a DWM of dimension \( D \), spatial sampling distance \( d \) and speed of sound \( c \). Ultimately \( f_{\text{update}} \) dictates the audio output quality from a DWM with large sample rates requiring denser meshes, more memory and taking longer to run, limiting even the most efficient large-scale implementations to offline generation only.

THE DYNAMIC DIGITAL WAVEGUIDE MESH

A DWM simulation is based on defining a spatio-temporal sampling grid whose geometry approximates the physical geometry of the object being modelled. A dynamic DWM implies that aspects of this geometry are subject to change, such that the DWM gets smaller or larger according to how the simulated object might be manipulated according to user input or external influence. The obvious method is to add or remove mesh points, updating scattering equations accordingly, on an iterative time-step basis, thereby removing aspects of the main system geometry or adding to it. This dynamic restructuring of waveguide elements and junctions at run-time can be problematic in maintaining the continuity laws governing the underlying scattering equations. Such changes can be minimized, helping to maintain stability, by defining new pressure components to be set to zero and disconnected pressure components to be disregarded. Mesh changes also generally occur at a slower rate than the actual mesh update frequency helping to further ensure a stable system. However, discontinuities appear in the audio output and are also clearly audible, limiting the possibilities for this potential solution [5].

Most DWM implementations result in a homogeneous structure which implies that the admittance terms in (Eq. 2) cancel, simplifying the scattering equations used. A signal will therefore propagate and scatter according to these lossless conditions until it interacts with a boundary where a reflection factor, determined by a relative change in admittance, defines reflective/absorptive behaviour [7]. However, in 1-D digital waveguide simulations an impedance change at a point within the system is often used to implement a change in the system’s physical characteristics. The classic example is the Kelly-Lochbaum vocal tract model, which is realised as a series of concatenated piecewise continuous tubes of different cross-sectional area according to tract shape. Changes in cross-sectional area are implemented as a change in impedance, which is otherwise constant along the length of each tube section, simulated using a 1-D digital waveguide [3]. This approach can be used for articulatory speech synthesis, and indicates that digital waveguide impedance variation is a potential solution for a dynamic DWM using (Eq. 2) directly with each admittance value \( Y \), being an additional dynamic variable now active and incorporated in the associated scattering junction equations at every timestep. Note that in accordance with, for instance, [3] and [5] impedance \( Z = 1/Y \) will be the dynamic variable discussed rather than \( Y \). The first step in defining a
dynamic system is determining the maximum mesh dimensions, and hence the total number of scattering junctions required for the geometry of the object being simulated, and the nature of the interaction required. For instance, if the object to be modelled is a 2-D rectangular membrane, the largest length/width should be determined together with the number of scattering junctions required along each dimension. With this base geometry in place, a default lowest impedance value $Z_{\text{min}}$ is set for each waveguide element as well as appropriate boundary terminations for the required simulation. An impedance map for the defined system is then calculated, determining how the impedance at each scattering junction/waveguide element will vary according to user input. This map is defined on a case-by-case basis according to the geometry, together with the nature of the simulation and interaction or dynamic variation required. Three such examples are presented in the following results.

RESULTS

This section presents three case studies where the dynamic DWM has been successfully used for sound synthesis. The first demonstrates a dynamically varying vocal tract simulation, validating the dynamic DWM approach using the system for which it was first proposed. The second case study explores how this method might be extended to general 2-D membranes. The third example returns to the simulation of the vocal tract, and explores how the dynamic DWM might be better controlled to facilitate improved articulatory speech synthesis.

The 2-D dynamic vocal tract model

![Diagram of 2-D dynamic vocal tract model](a) Raised cosine impedance function $Z(x,y)$ for the 2-D dynamic DWM vocal tract model; (b) vocal tract area function information; (c) resulting DWM impedance map.

A constant width 2-D rectangular 4-port DWM, 17.5 cm long, $f_{\text{update}} = 44.1\text{kHz}$, is used as the basis for a 2-D vocal tract simulation. In the Kelly-Lochbaum model, tract shape is defined as a series of concatenated cylindrical acoustic tubes, of different cross-sectional area $A$ and impedance, $Z$. With the 2-D model, area function information $A(x)$ is used to construct an impedance map along the length, $x$, of the 2-D mesh. The waveguide element impedance across the mesh width, $y$, is then varied according to each $A(x)$. A minimum impedance channel $Z_{\text{min}}$ is defined as the lowest value across the range of vowels to be simulated, corresponding to the largest value of $A(x)$ and so determining a maximum tract width opening, $w$. An impedance map is constructed for a particular vowel tract shape such that $A(x)$ along the length of the tract walls corresponds to a maximum impedance value $Z_x$. An impedance curve varying from $Z_{\text{x}}$ to $Z_{\text{min}}$ and back to $Z_x$ at the opposite wall is then defined across the tract width $y$ according to a raised cosine function, with $Z_{\text{min}}$ equidistant between tract walls and as shown in Fig. 2(a) and given in (Eq. 4):

$$Z(x,y) = Z_x - \left(\frac{Z_x - Z_{\text{min}}}{2}\right) \left[1 + \cos \left(\frac{\pi}{w} \left(\frac{y}{w} - \frac{1}{2}\right)\right)\right]$$  \text{(Eq. 4)}

Fig. 2(b) shows typical cross-sectional area function information $A(x)$ as it varies along the length of the vocal tract from glottis to lips. Fig. 2(c) is the corresponding impedance map generated according to this information and (Eq. 4) imposed across and along the underlying rectangular 2-D DWM. Areas of higher impedance are represented by lighter shading and the $Z_{\text{min}}$ impedance channel can be observed as the darker area along the centre of the map. The result of this method is the possibility for dynamic articulation as shown in Fig. 3 demonstrating smooth interpolation between area function data for the /u/ - ‘food’, and /Λ/ - ‘but’, vowels.
The 2-D dynamic membrane

A rectangular membrane of dimensions A (2.86m x 3.3m) is simulated using a 6-port DWM with \( t_{\text{update}} = 44.1 \text{kHz} \). Over 80,000 time-steps this membrane is smoothly reduced to one of size B (1.54m x 1.33m) by increasing \( Z \) using the linearly varying impedance map:

\[
Z(l) = \begin{cases} 
Z_{\text{min}}, & Z_l - l \cdot m < Z_{\text{min}} \\
Z_l - l \cdot m, & Z_l - l \cdot m \geq Z_{\text{min}} 
\end{cases}
\]  
(Eq. 5)

Where \( l \) is defined as a layer of the mesh membrane varying from the outer boundary edge (\( l = 0 \)) at A to the final target size B (\( l = N \)) and \( m \) is the gradient of the impedance change. The first five theoretical modal frequencies for mesh A are (1,0) = 52.0Hz; (0,1) = 60.0Hz; (1,1) = 79.4Hz; (2,0) = 104Hz; (0,2) = 120Hz. For B these become (1,0) = 111Hz; (0,1) = 129Hz; (1,1) = 170Hz; (2,0) = 223Hz and (0,2) = 258Hz. Screenshots from the simulation are shown in Fig. 4(a) where the z-axis denotes increasing impedance. The resulting smooth change in modal frequencies is highlighted in the spectrogram in Fig. 4(b).

Other examples demonstrating the possibility for novel synthesis or interaction are also possible. Fig. 4(c) shows the same membrane (of dimensions A) where high impedance regions based on a variation of the raised cosine impedance function defined in (Eq. 4) are slowly introduced and then removed over the length of the simulation.

Articulatory Vocal Tract Speech/Singing Synthesis

A dynamic real-time synthesis engine of this nature raises questions of how best the user might interact with such a model, particularly if the synthesized object is abstract in nature as in the
example shown in Fig. 5. Of particular interest is how a multi-parametric vocal synthesis system based on a 2-D DWM might be better articulated to give more natural speech output. The 2-D vocal tract system has therefore been further adapted to import $A(x)$ data as a series of text files, with dynamic interpolation from one file to the next allowing more complex articulation than the dipthong synthesis shown in Fig. 3. The area function information is generated by the APEX system [8], a tool that can be used to synthesize sound and generate articulatory voice related parameters, based on the positioning of lips, tongue tip, tongue body, jaw opening and larynx height, all mapped from X-ray data. The phrase “A Boy I Adore” is synthesized as a series of nine vocal tract profiles, /a/-/b/-/O:/-/i/-/a:/-/i/-/a/-/d/-/o/, with a vowel transition time of 250ms, the results of which are shown in Fig. 5. The changes in formant pattern from vowel to vowel are clearly evident, with the tract constrictions for /b/ and /d/ being particularly noticeable, demonstrating the potential for high-level synthesis using a dynamically varying DWM.

CONCLUSIONS

This paper has introduced how the 2-D DWM might be extended to facilitate dynamic synthesis and interaction based on variable impedance calculations at each scattering junction. Although already shown as appropriate for vocal tract modelling, this work has explored improved articulation using a third party application to generate the required synthesis parameters, resulting in the generation of simple phrases. This approach has been extended to dynamically varying membranes, demonstrating its potential for wider sound synthesis. Further research will explore the implementation of these methods as part of a wider synthesis framework, thereby facilitating improved user interaction and control of the underlying models. Sound examples and a demonstration version of the vocal tract synthesis system are available online at [9].

References: