An Evaluation of Rectilinear Digital Waveguide Mesh in Modelling Branch Tube for English Nasal Synthesis

**Anocha Rugchatjaroen, PhD**NECTEC, National Science and Technology Development Agency (NSTDA),   
Patumtani, 12120, Thailand

**Prof David M Howard**

Department of Electronic Engineering,

Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom

**Abstract**: This research presents the evaluation from applying a Two-Dimensional Rectilinear Digital Waveguide Mesh (2-D RDWM) to model the vocal and nasal tracts for synthesizing the three English nasal consonants /m/, /n/ and /ŋ/. The rectilinear meshing represents the propagated pressure according to the cross-sectional area at each tract position. Nasal tract coupling is achieved by modifying the number of ports from four to five at the location of the velopharyngeal mesh junction to attach nasal tract mesh. The proposed system exhibits zeros from the branch tube (closed-end oral cavity) and also nasal frequencies from the nasal tract. Spectrograms of the acoustic output for synthesized nasals exhibit zeros and nasal tract resonant frequencies comparable to those found in synthesized and real speech.

Index Terms: speech modelling, English nasal, 2-D Digital Waveguide Mesh

# Introduction

In the majority of articulatory tube modelling for human speech synthesis, the vocal tract is viewed as a structure consisting of small tubes with corresponding cross-sectional areas that are used to represent the vocal tract characteristics in the acoustic model. Each cross-sectional area can be modelled as an electrical transmission line and these can be combined to simulate the shape of the vocal tract, being changed over time to model vocal tract articulation during speech. The acoustic behaviour of each individual tube is described in terms of a volume velocity *u(x,t)* and a pressure *p(x,t)* deviation. For example, in Kelly-Lochbaum vocal tract modelling [1], the pressure is decomposed into the sum of a leftward-travelling pressure wave *pl* and a rightward-travelling pressure wave *pr*, as is the volume velocity which is similarly a sum of leftward- and rightward-travelling velocity waves (*ul* and *ur*). The travelling pressure and volume velocity waves are related to each other via the tube admittance or impedance [2]. Alternative articulatory and acoustic modelling methods have been proposed [3-7].

From the point of view of computational modelling, the computational fluid dynamics is well-developed and -used to replicate the air-flow behaviors in such a narrow and complex space as vocal tract. In 2008, Doorly gave a brief discussion on pros and cons from using such computational modelling for medical studies [8]. From the view of physical modelling, Smith proposed the model of motion in the medium after the wave equation using digital filter design and scattering theory in the mid-1980s, which offered significantly lower computational expenses than other sound synthesis methods, using two directional delay lines to simulate the two travelling non-interacting waves. Stefan Bilbao suggested and discussed the relationship between digital waveguides and standard time domain numerical methods such as FDTD in [2, 9, 10]

From an acoustic modelling perspective, the digital waveguide mesh (DWM), used in a musical or room acoustic modelling, was introduced to model vocal tract acoustic behaviour by Cook in 1990. He used it to model a vocal tract in his singing synthesizer. He considered the spatial variation of the cross-sectional area of the vocal tract as the multiple waveguide segments separated by scattering junctions [11]. Later on, Mullen applied a two-dimensional (2-D) mesh structure by using tract area functions and linear control over formant bandwidths across a range of vowels [12, 13]. This paper develops work with this 2-D DWM through the addition of a nasal cavity.

For a nasal, Ladefoged [14] and Stevens [15] explained the relevant articulatory phonology and Johnson in chapter 9 informatively describes its acoustic phonetics [16]. Cox also explained that "The most general effect of adding nasal resonance to oral resonance is an overall loss of power." [17]. Previous research in nasal synthesis has been carried out by [18, 19, 20], and in nasality and nasalization including their acoustic properties in vowel nasalization by [21-26]. In this article, we describe the synthesis of English nasal consonants using a 2-D RDWM, which has the potential to require less computation compared to other physical modelling techniques.

This article is organized as follows. In the next section, we describe the main physiological characteristics of English nasals and their acoustic behaviour. Section 3 explains our proposed method and Section 4 presents experiments and results. Section 5 comprises the conclusions.

# English nasal resonance

Nasalization involves the coupling of the nasal cavity to the oral tract during speech production. To achieve this, a speaker lowers his/her velum to enable air to flow through the nasal cavity. In 2013, Stevens and Hanson included a short easy-to-understand summary of articulatory events of nasal consonants in *Theories and Models of Speech Production* [27].

Several studies have been published on nasal vowels, especially on French nasal vowels. With image technology, images of articulation are involved and sizes of the velopharyngeal opening (VPO) are studied to research their acoustic effects, such as effects on F1, F2, F3, and sometimes F4 and also their transitions, etc. [21, 22], [24-26], [28].

In English, there are three nasal consonants – /m/, /n/ and /ŋ/; their production involves complete closure with bilabial, alveolar or velar contact respectively, and air exits via the nasal cavity only. Figure 1 shows the positions of the articulators as magnetic resonance images (MRI) taken at the York Neuroimaging Centre (YNiC), (General Electric 3.0 T HDx Excite Scanner) during the production of /m/ (left), /n/ (mid) and /ŋ/ (right). The white arrows show the widely opened velopharyngeal port. White circles indicate the place of articulation of each nasal consonant.

In acoustic phonetic theory, nasals have acoustic characteristics that are related to the resonances in the nasal cavity which cannot be changed in volume; these are (3 x 350.2) / (4 x 0.22) from *(2n-1)c/4L* (where *c* is the velocity of sound and *L* is the tube length), giving 1,193 Hz, 1,989 Hz and 2,785 Hz for a nasal tract of length 0.22 m. The second formant (1,193 Hz in the example) has the highest amplitude in the spectrogram. The other key characteristic acoustic property of nasal sounds is their anti-formants: the resonances of the side-branch oral cavity. The centre frequencies of these depend on the length of the oral cavity in each of the three nasals. For instance, the mouth cavity in /m/ is approximately 0.08 m long, then the first two quarter-wave resonant frequencies are at 1,100 Hz and 3,300 Hz, and therefore an anti-formant around 1,100 Hz is anticipated.



Figure 1: Cross-sectional MRI from a speaker in Speed [29] holding his articulators for pronouncing /m/(left), /n/(mid) and /ŋ/(right).

# 2-D digital waveguide mesh for nasals

In 1-D DWM for vocal tract modelling, the vocal tract length *l* is decomposed into a series of cylindrical segments of length *h* such that *N* = *1/h* when *N* is the number of segments. The scattering junction *J* is interconnected with pairs of digital unit sample delay lines. The pressure and volume velocity at position *x* at time *t* are the composition of their left- and right travelling components [1].

*p*(*x*, *t*) = *p*(+) + *p*(-) and *u*(*x*, *t*) = *u*(+) + *u*(-) (1)

From the 1-D scattering interpretation which was proposed by Kelly and Lochbaum in 1962, twenty-eight years later Cook successfully implemented his singing voice synthesizer [11]. Duyne and Smith proposed an extension to 2-D DWM in 1993 for their physical modelling for sound synthesis [30], the model was also implemented into an application, a dynamic vowel synthesizer, using 2-D DWM by Mullen in 2007 [31]. In 2-D, a scattering junction must have three or more ports in meshing. The difference in the number of ports causes different meshing topology. In 2005, Compos and Howard explained the details of the scattering characteristics, and the pros and cons of each topology [32]. Here, we employ the rectilinear mesh for the ease of mapping junction positions to real tract shape.

Figure 2 shows an example structure of meshing a 2-D rectilinear DWM fitting in a vocal tract tube model. A black square represents a rectilinear scattering junction. Each pair of parallel lines represents a digital unit sample which is strung together with others at junctions to create a mesh. Note that tube diameters are used as weight function through the y-axis, not the structure.

Figure 2: An example of a 2-D rectilinear DWM with 4-port 90° connected junctions for the illustrated vocal tract tube model.

At the junction the *N* incoming and outgoing pressures are summed and distributed at the same temporal and sampling location. To calculate the pressures at the junction, all incoming pressures are summarized and then appropriately weighted, depending on the type of topology. For a rectilinear topology, the pressure at the junction is

(2)

where *pJ* is the pressure at the junction and *pJ,i*+ is the incoming pressure from each port *i* [31].

The pressure is propagated through the mesh junctions. At each waveguide (*i*) output pressure from each junction (*J*) is calculated by

(3)

The incoming pressure is the output pressure from a neighbouring junction at the previous time step, and an outgoing pressure becomes the incoming pressure of a neighbouring junction at the next time step.

At the junctions themselves, the acoustic impedance *Z* represents the ratio of the pressure relationships, which is constant in an unbounded homogenous medium. Sometimes *Z* is replaced by the equivalent acoustic admittance *Y* for convenience. This impedance or admittance is used in 1-D models to represent the relationship between the area functions of two discontinuous tubes. In a static rectangular 2-D DWM, the changing of the area function would have a small effect in the cross-sectional plane. A constriction could be applied as raised impedance regions in order to encourage cross-tract reflections.

The raised cosine impedance function in Mullen, 2007 is also adopted and used in this research. Its aim is to define the varying impedance across the tract as a bell-shaped curve. This curve represents the ‘impedance hill’ across the tract, with the highest impedance being at the borders and the lowest at the midline. With Cartesian coordinates, the impedances are calculated by equation (4) which is dependent on the y position. The maximum impedance values are at *y=0* and *y=n* and the minimum is at *y=n/2*. It is like the half-circular contour of admittance parameter *Zx* which is located at the point of constriction *x*. Therefore along the y-axis, *Zx,0 … n* represents the impedance values at different *y* positions; the maximum impedance value of a constriction *x* is *Zx,0* and *Zx,n* and the minimum is *Zx,n/2* or *Zmin* which is equal to *Ztube*. A raised cosine smoothing is used to control the impedance hill [31], where *Z(x,y)* is then defined as

(4)

where *w* is the tract width corresponding to cross-sectional area function at the position *x*. In this research, the highest impedance is set and used (called *Zstop*). It is the impedance for the modelling of a complete cut-off of air-flow which is used in place of closure in the oral tract in this research. With this impedance control and the attachment of a nasal tract, nasals can be modelled.

A simulation of wave propagation in the oral tract [31] has been modified by adding a nasal tract to create a more complete overall vocal tract model. This enables pressure to propagate through the port (bi-directional from the oral tract to the nasal tract, and from the nasal to the oral tract). Figure 3 shows an example structure of the construction of a 2-D DWM with an oral and nasal tract coupled along the x-axis.

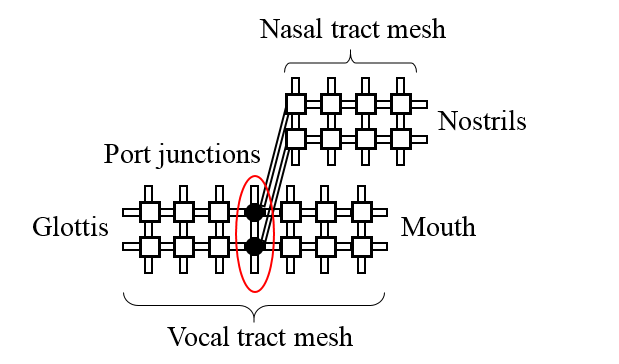


Figure 3: Vocal tract rectilinear 2-D mesh structure with an oral and nasal tract coupled along the x-axis.

At the connecting junctions, the cosine impedance is also used here. The degree of velopharyngeal opening parameter weights the first cross-sectional area of the nasal tract (counting from velopharyngeal port to nostrils), then the weighted area is converted into an impedance and used as *Zmin*, and then they are weighted with the cosine function and spread into each waveguide units in Y by equation (4). The size and number of waveguides are calculated from the frequency-dependent dispersion equation from [33].

(4)

where *c* is the speed of sound, *N* is the number of dimensions of the waveguide mesh, *d* is the waveguide length and *fs* is the sampling frequency.

In our simulation, *fs* is 22.050 kHz, *c* is 350.2 m/s, *N* is 2 then *d* is 0. 022460 m (2.24 cm). The maximum vocal tract shape from the MR images is 0.175 m x 0.05 m, so the simulation needs 8 x 3 waveguides, containing 14 junctions, while the nasal tract shape is 0.10 m x 0.05 m, and therefore needs 5 x 2 waveguides, containing 8 junctions (the structure is similar to that in Figure 3).

# Experiments and results

To study the nasalization of the proposed system, the first set of four vowels (/a/, /ә/, /i/, /u/) were selected, and these were synthesized in both nasalized and non-nasalized forms. The tract shapes were adopted from [31]. Figure 4 shows the frequency spectra for the synthetic nasalized and non-nasalized versions of these four vowels, overlaid to enable direct comparison. The area functions used for the vowels were adopted from extracted area functions from MR images in Mullen, 2006. An FFT of results in Figure 4 shows the existence of the second formant of nasal murmur at around 1 kHz in all results for the nasals which can be predicted by (3 x 350.2) / (4 x 0.22), *(2n-1)c / 4L*, which equals 1,193 Hz.

Based on the nasal murmur characteristic, which is nasal formants relating to the nasal plus pharyngeal cavity, the F1 of nasal murmur should have a higher amplitude followed by lower amplitude caused by damping [16]. Cox also confirmed [17]. Kent and Read drew the idealized damped amplitude of nasal vowels compared to non-nasalized vowels in Figure 5-34 in [34]. However, 2-D DWM is used to propagate pressure only, not absorption. Managing the wall reflection coefficient as in [35] results in changes in resonance bandwidths, but not sufficient to account for observed nasal damping in natural speech. Nonetheless, we can see the second formant of nasal murmur at around 1 kHz in all results for the nasals which can be predicted by (3 x 350.2) / (4 x 0.22), *((2* x *2)-1)c / 4L*, which equals 1,193 Hz.

Figure 4:FFT of nasalized (grey line) and non-nasalized (dark line) versions of four 22.050 kHz sampling rate synthesized vowels: /a/ (top left), /ә/ (top right), /i/ (bottom left) and /u/ (bottom right).

Note that our results also support Childers’ review of Meada’s statement that the high vowels such as /i/ and /u/ are more nasalized than the mid and open vowels; an effect that is borne out in these results [36].

The position of articulators and vocal tract shape of three images in Figure 1 (/m/, /n/ and /ŋ/) were extracted and used in this research. Here we assume that the coupled nasal tract includes all the tissues and sinus spaces, hence the resulting 3-D segmented nasal tract shape was a large balloon-shape connected to the velopharyngeal port (two nostrils were merged together as one single open-end). Then the 3-D tract shape was passed to the bisection method to find cross-sectional area functions and tract lengths.

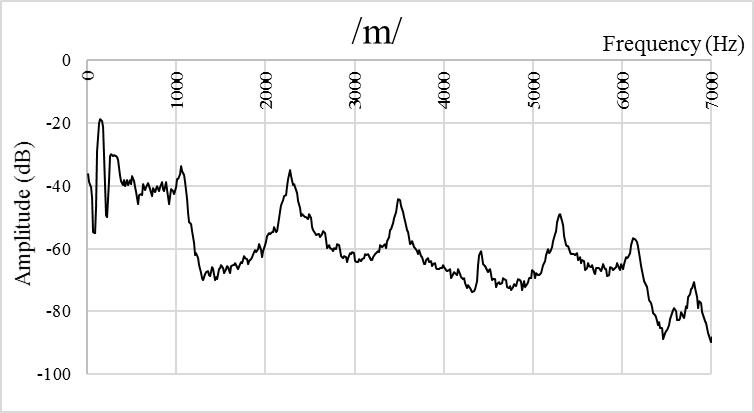
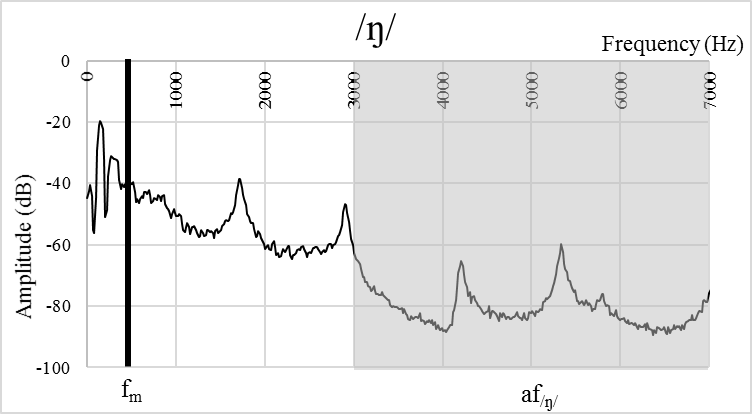
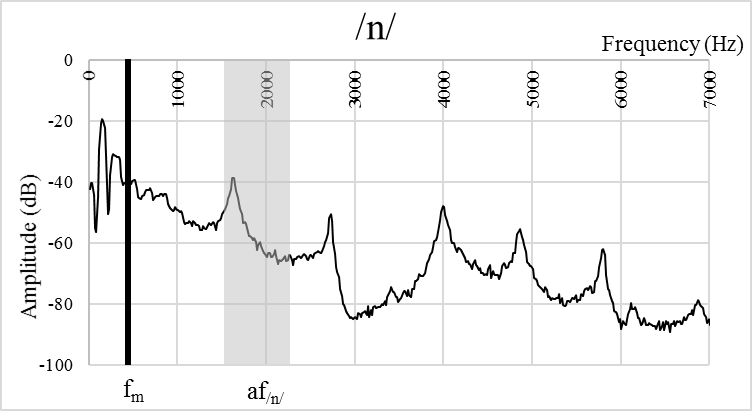
In detail, the cross-sectional areas were extracted from the 3-D data using *vtk* tools as described in section 6.3 of [29]. The step starts with creating a line connecting the centre of the glottis and the mouth and then finding a middle point (to be used as a reference point). Then a perpendicular plane that has normal vectors directing to the centre of the glottis and mouth and cut past the middle point is created and used as the first cutting-plane. Finally, the cross-sectional area which is found on the plane is extracted. These bisectional steps are then repeated recursively by referencing to the new centroid of new cross-sectional areas until the target number is reached [27].

One of the main acoustic characteristics of nasal consonants, the nasal anti-formant (zero), can be observed in Figure 5, which shows long-term spectra of synthesized versions of /m/, /n/ and /ŋ/ at a sampling frequency of 22.050 kHz (waveguide size 0.022m). The results show anti-formants at around 1,700 Hz in /m/ and /n/ and 3,060 Hz in /ŋ/. The similarity of the centre frequencies of the zeros in /m/ and /n/ is a result of the very close (<0.022 m) physical proximity of the closure points in /m/ and /n/, which requires the model to be constricted at the same waveguides (note that there are slightly different y-axis neighbouring waveguide sizes that cause slightly different zeros).

Figure 5: Long-term analysis of synthesized /m/, /n/ and /ŋ/ from 22 kHz of sampling frequency.

These results were synthesized using a 22.050 kHz sampling rate, which equates to a waveguide length of 0.022 m. The oral cavity for schwa is 0.175 m in length in these vocal tract simulations, which means that it is simulated by the last three waveguides (0.022 x 3 = 0.066 m). (Note that width is not discussed here, because cross-mode resonance only becomes relevant at higher frequencies and our resonance considerations are lengthwise only.) Of these three waveguide elements, the last represents the pressure propagation at the lips, the penultimate approximately represents the post-alveolar phonetic place of articulation, and the one before that approximately represents the palatal phonetic place of articulation; /ŋ/ requires a velar constriction, /m/ a bilabial constriction and /n/ an alveolar constriction.

Figure 6 shows results from using a higher number of junctions by synthesizing at a considerably higher sampling rate (176.4 kHz). With a size of 0.175 x 0.05 m2 for vocal tract and 0.11 x 0.05 m2 for nasal tract, a mesh for synthesizing output at 176 kHz needs a total number of 1,054 + 663 = 1,717 junctions. The synthesized voices show significant nasal characteristics with a significant anti-formant for /m/ at 1,300 and 6,400 Hz, for /n/ at around 3,000 Hz and for /ŋ/ at around 4,000 Hz which are close to the theoretical values (see Table 1). Reference point at fm and range of af/m/, af/n/and af/ŋ/ are provided of the approximated murmur frequency, and range of anti-formant of /m/, /n/ and /ŋ/, respectively. Moreover, Table 1 shows a comparison of the anti-formants in the analysis in Figure 6 which can be compared to those in [21] and [37].



fm

af/m/

Figure 6: Long-term analyses of synthesized /m/, /n/ and /ŋ/ with a sampling frequency of 176.4 kHz.

oward 2009) and our synthesized speech..n of the aniformants of Table 1: A comparison between the anti-formants from theoretical [37], observed [21] and our synthesized speech.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Theoretical Prediction  [26] | Phonetician Observation [21] | Synthesized Voice |
| /m/ | 1,000 Hz | 750 - 1,250 Hz | 1,300 and 6,400 Hz |
| /n/ | 3,500 Hz | 1,450 – 2,200 Hz | 3,000 Hz |
| /ŋ/ | 5,000 Hz | 3,000 Hz and above | 4,000 Hz |

A listening test was carried out to evaluate the perception of the nasalization synthesis in terms of whether changes were perceptible to trained phonetician listeners. The experiment needed trained phoneticians because there was no real recorded speech provided to allow participants to compare sounds; hence the participants had to have the ability to pick up and perceive vowels and nasalization themselves. Moreover, they had to rate naturalness for /m/, /n/ and /ŋ/ without referencing sounds, and therefore their expertise was strongly needed.

The test contained two parts. The first part was designed to test the nasalization of the proposed system. Pairs of nasalized and non-nasalized tokens were played and participants were asked to select the nasalized version. This experiment was set under the purpose of testing damping and zero frequency perception, hence four non-nasalized/nasalized vowel pairs (/a/, /ã/), (/i/, /ĩ/), (/u/, /ũ/) and (/ə/, /ә̃/) were contained in this test. Eight sounds were randomly put into a listening table given to participants, together with an answer sheet. In the answer sheet there were check boxes for the testing phone set and a blank space for giving the phonetic symbol if they thought the sound was not recognized as provided in the answer sheet.

Eleven trained phoneticians performed the test in an office environment, perceiving sounds from the same pair of headphones. They were of an average age of 27 years ranging from 24 to 38 years. Nine of them were female while two of them were male. Half of them are English native speakers and the rest are either doing or obtaining a degree of Linguistics or Phonetics from an English environment school. The results from the first part of the test showing the number of tokens heard by participants as being nasalized or non-nasalized for each vowel stimulus are shown in Table 2. It can be seen that the synthesis of all vowels except /a/ and /ã/ provided sufficient acoustic cues to enable the nasalized/non-nasalized difference to be accurately perceived in all cases. Clearly, there are issues with the synthesis of /a/, but there is evidence to suggest that the majority of the nasalized /ã/ tokens were heard as being nasalized. Rather surprising is the fact that the non-nasalized /a/ vowel was perceived as being nasalized for the majority of its tokens.

oward 2009) and our synthesized speech..n of the aniformants of Table 2: Results from the nasalized/non-nasalized vowel listening test, indicating the number of participants (of 11) perceiving non-nasalized or nasalized for each vowel.

|  | Heard as non-nasalized | Heard as nasalized |
| --- | --- | --- |
| /a/ | 8 | 3 |
| /ã/ | 3 | 8 |
| /i/ | 11 | 0 |
| /ĩ/ | 0 | 11 |
| /u/ | 11 | 0 |
| /ũ/ | 0 | 11 |
| /ə/ | 11 | 0 |
| /ә̃/ | 0 | 11 |

The second part of the test was designed to evaluate the synthetic English nasals /m/, /n/ and /ŋ/; participants were asked to label these and give a naturalness percentage, indicating how similar they believed these were to the natural versions. Again, two copies of the three synthesized nasal consonants were put randomly into a listening table given to participants, together with the answer sheet.

Figure 7 shows the results of the listening test for the 22.050 kHz synthesized nasal consonants (/m/, /n/ and /ŋ/) which were tested subjectively in the second part.

Figure 7: Scores from labelling /m/, /n/ and /ŋ/ counted by number of participants who had given their mark.

As shown in the figure, the synthesis of the bilabial nasal /m/ appears fairly successful, but the results suggest that there is marked perceptual ambiguity in the perception of the synthetic results for /n/ and /ŋ/. Synthetic /ŋ/ was mostly marked as /n/ while synthetic /n/ was confusing and marked as /m/ or /ŋ/. These are the results of mismatched zeros which are shown in Figure 5. The synthetic /n/ has zero at around 1,700 Hz while it contains another of much less energy at around 4,600 Hz which are in the range of zeros for /m/ and /ŋ/, respectively, while synthetic /ŋ/ has a zero at around 3,000 Hz which is almost in the range of zeros for /n/. Therefore, only /m/ was perceived correctly when the sounds were synthesized at 22 kHz of sampling frequency.

For the group of answers that marked the phone as we proposed, their opinions in percentage of similarity between the synthesized phones and their possible natural versions were counted and averaged as 37%, 35% and 30% for /m/, /n/ and /ŋ/, respectively. Overall then, there is still room in perceptual confusing reduction between the nasal consonants and the confidence with which the sounds are being labelled to be improve in future work.

# Conclusions

This work has implemented oral and nasal tract modelling using a 2-D RDWM as a dual branch tube to model the vocal tract. At the velopharyngeal port where the oral and nasal tracts connect, the structure was modified from having 4-port to 5-port junctions. The number of ports in the wave scattering equation was modified and applied differently at the corresponding junction, in order to add the possibility of synthesizing nasal frequencies and zeros to the model. The proposed model was tested by synthesizing four nasalized vowels and English nasal consonants.

The main acoustic characteristics of nasals – low-frequency nasal cavity murmur and the main zero related to the coupled oral tract – are identified in the spectrogram of the output sounds. Objectively compared with results from the well-established Praat [27], the frequency responses from both systems are still not exactly in the range observed by phonetician (shown in Table 1) such as closing mouth for /m/ in Praat resonates the first zero at around 1,500 Hz, etc. More accurate results from both systems can be achieved by controlling more precise place of articulation. The subjective results depict the accuracy of human perception while the different centre frequencies of the zeros in 22.050 kHz sampled results are likely to be responsible for the confused perception.

However, the objective and subjective test results show promise for the synthesized speech from the 2-D waveguide mesh. The perceptual confusion in the last set of results was the result of a too low frequency synthesis. This shows that we have to trade between computation cost and synthetic accuracy and that 22.050 kHz is not sufficiently high. This indicates some room for improvement and further studies could, for example, synthesize the sounds in a different mesh topology or in a higher dimension, use a 3-D mesh [29] or divide the nostril into two and study the effect, or study the effect of the degree of velum opening for nasal vowels or add damping control in nasal murmurs which at the moment there is no parameter to control.

# Acknowledgments

I would like to express my gratitude to Dr Christin Kirchhübel and Sukanya Thaitechawat for all their help and suggestions and to Dr Matthew Speed for providing the English nasal MRI vocal and nasal tract images used in this research.

# Rreferences

|  |  |
| --- | --- |
| [1] | J. L. Kelly and C. C. Lochbaum, "Speech Synthesis," in *the Fourth International Congress on Acoustics*, Copenhagen, Denmark, 1962. |
| [2] | S. Bilbao, "1-Introduction," in *Wave and Scattering Methods for Numberical Simulation*, Chichester, J. Wiley, 2004. |
| [3] | P. B. T. M. P. Rubin, "An articulatory synthesizer for perceptual research," *Journal of the Acoustical Society of America,* vol. 70, pp. 321-328, 1981. |
| [4] | C. P. Brownman and L. GoldStein, "Towards and articulatory phonology," in *Phonology Yearbook 3*, 1986, pp. 219-252. |
| [5] | P. Boersma, "Functional Phonology," Holland Academic Graphics, Netherlands, 1998. |
| [6] | P. Birkholz, "3D-Artikulatorische Sprachsynthese," Logos-Verlag, 2005. |
| [7] | P. Šidlof and S. Zörner, "Computational aeroacoustics of human phonation," in *Experimental Fluid Mechanics 2012*, Czech Republic, 2013. |
| [8] | D. J. Doorly, D. J. Taylor and R. C. Schroter, "Mechanics of ariflow in the human nasal airways," *Respiratory Physiology & Neurobiology,* vol. 163, pp. 100-110, 2008. |
| [9] | J. O. Smith III, "On the equivalence of digital waveguides and finite difference time domain schemes," 2004. [Online]. Available: http://arxiv.org/abs/physics/0407032/. |
| [10] | M. Karjalainen and C. Erkut, "Digital waveguides vs. finite difference schemes: Equivalence and mixed modeling," *EURASIP Journal on Applied Signal ,* pp. 978-989, 2004. |
| [11] | P. Cook, "Identification of Control Parameters in an Articulatory Vocal Tract Model with Applications to the Synthesis of Singings," PhD Thesis, the Department of Electrical Engineering, Standford University, 1990. |
| [12] | J. Mullen, D. M. Howard and D. T. Murphy, "Digital Waveguide mesh modeling of the vocal tract acoustics," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2003. |
| [13] | J. Mullen, D. M. Howard and D. T. Murphy, "Waveguide Physical Modeling of Vocal Tract Acoustics: Flexible formant bandwidth control for increased model dimensionality," *IEEE Trans. Speech Audio Processing,* vol. 14, no. 3, pp. 964-971, 2006. |
| [14] | P. Ladeforged, Vowels and Consonants An Introduction to the sounds of languages, Oxford: Blackwell, 2001. |
| [15] | K. N. Stevens, Acoustic Phonetics, Cambridge, Massachusetts: The MIT Press, 1998. |
| [16] | K. Johnson, Acoustic and Auditory Phonetics (Third Edition), Chicester: John Wiley & Sons, 2012. |
| [17] | F. Cox, "Consonant Acoustics: The Acoustic Characteristics of Nasals," Macquarie University, 2008. [Online]. Available: http://clas.mq.edu.au/speech/acoustics/consonants/nasalweb.html. [Accessed 09 06 2014]. |
| [18] | A. I. Russell, "Synthesis of Nasal Consonants: A Theoretically Based Approach," Master Thesis, Computer Science and Engineering, the Massacusetts Institute of Technology, 1999. |
| [19] | J. G. Nieveen, "Modeling the Liquid, Nasal, AND Vowel Transitions OF North American English Using Linear Predictive Filters and Line Spectral," Master Thesis, the Department of Electrical Engineering, Utah State University, 2013. |
| [20] | A. Patki and D. O. Frank-Ito, "Characterizing Human Nasal Airflow Physiologic Variables by Nasal Index," *Respiratory Physiology & Neurobiology,* p. 66–74, 2016. |
| [21] | O. Fujimura, "Analysis of nasal consonants," *Journal of the Acoustical Society of America,* pp. 1865-1875, 1962. |
| [22] | S. Maeda, "Acoustic Correlates of Vowel Nasalization: A Simulation Study," *Journal of the Acoustical Society of America,* 1982. |
| [23] | S. a. S. K. N. Hawkins, "Acoustic and Perceptual Correlates of the Non-Nasal-Nasal Distinction for vowels," *Journal of the Acoustical Society of America,* pp. 1560-1575, 1985. |
| [24] | K. N. F. a. H. S. Stevens, "Some Acoustical and Perceptual Correlates of Nasal Vowels," in *Festchrift fur Ilse Lehiste*, Dordrecht, Holland, 1987, pp. 241-254. |
| [25] | S. Maeda, "Acoustics of Vowel Nasalization and Articulatory Shift in French Nasal Vowels," in *Nasals, Nasalization, and the Velum*, New York, 1993, pp. 147-167. |
| [26] | G. Feng and E. Castelli, "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization," *J. Acoust. Soc. Am.,* pp. 3694-3706, 1996. |
| [27] | W. J. Hardcastle, J. Laver and F. E. Gibbon, The Handbook of Phonetic Sciences Second Edition, Oxford, UK: Wiley-Blackwell, 2013. |
| [28] | D. Y. Wang, P. L. Heow and G. R. Bruce, "Impacts of Fluid Dynamics Simulation in Study of Nasal Airflow Physiology and Pathophysiology in Realistic Human Threedimensional Nose Models," *Clinical and Experimental Otorhinolaryngology,* pp. 181-187, 2012. |
| [29] | M. Speed, "Voice Synthesis using the Three-Dimensional Digital Waveguide Mesh.," PhD thesis, University of York, York, 2012. |
| [30] | S. V. Duyne and J. Smith, "Physical modelling with the 2-D Digital waveguide mesh," in *Computer Music Conf.*, Tokyo, Japan, 1993. |
| [31] | J. Mullen, "Physical Modelling of the Vocal Tract with the 2D Digital Waveguide Mesh," A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Electronics, University of York, York, 2006. |
| [32] | G. Campos and D. Howard, "On the computational efficiency of different waveguide mesh topologies for room acoustic simulation," *IEEE Trans, Speech Audio Processing,* vol. 13, no. 5, pp. 1063-1072, 2005. |
| [33] | U. Stephenson, "Comparison of the Mirror Image Source Method and the Sound Particle Simulation Method," *Applied Acoustics,* pp. 35-72, 1990. |
| [34] | R. D. Kent and C. Read, "Chapter 5: The acoustic characteristics of consonants," in *The acoustic analysis of speech - 2nd ed.*, NY, Singular Press, 2002. |
| [35] | A. Rugchatjaroen and D. M. Howard, "A Study on Dynamic Vocal Tract Shaping for Diphthong Simulation using A 2D Digital Waveguide Mesh," in *the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, York, UK, 2012. |
| [36] | D. Childers, Speech Processing and Synthesis Toolboxes, NY: John Wiley, 2000. |
| [37] | D. Howard and D. Murphy, Voice Science acoustics and recording, Oxford: Pural Publishing, 2008. |
| [38] | S. Lemmetty, "Review of Speech Synthesis Technology," Master's Thesis, Laboratory of Acoustics and Audio Signal Processing, Helsingki University of Technology, 1999. |
| [39] | S. Maeda, "The role of the sinus cavities in the production of nasal vowels," in *ICASSP*, 911-914, 1982. |
| [40] | T. Arai, "The Replication of Chiba and Kajiyama’s Mechanical Models of the Human Vocal Cavity," *Journal of the Phonetic Society of Japan,* vol. 5, no. 2, pp. 31-38, 2001. |
| [41] | S. Bilbao, Numerical Sound Synthesis Finite Difference Schemes and Simulation in Mulsical Acoustics, West Sussex: John Wiley & Sons, 2009. |
| [42]  [43] | P. Palo, "A Review of Articulatory Speech Synthesis," Master’s Thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, Helsinki, 2006.  " Human nose temperature too cold for avian influenza," Imperial College London, 15 05 2009 [Online]. Available: http://www.news-medical.net/news/2009/05/15/Human-nose-temperature-too-cold-for-avian-influenza.aspx [Accessed 22 01 2017]. |

# List of Figure Captions

Figure 1: Cross-sectional MRI from a speaker in Speed [29] holding his articulators for pronouncing /m/(left), /n/(mid) and /ŋ/(right).

Figure 2: An example of a 2-D rectilinear DWM with 4-port 90° connected junctions for the illustrated vocal tract tube model.

Figure 3: Vocal tract rectilinear 2-D mesh structure with an oral and nasal tract coupled along the X-axis.

Figure 4: FFT of nasalized (grey line) and non-nasalized (dark line) versions of four 22.050 kHz sampling rate synthesized vowels: /a/ (top left), /ә/ (top right), /i/ (bottom left) and /u/ (bottom right).

Figure 5: Long-term analysis of synthesized /m/, /n/ and /ŋ/ from 22 kHz of sampling frequency.

Figure 6: Long-term analyses of synthesized /m/, /n/ and /ŋ/ with a sampling frequency of 176.4 kHz.

Figure 7: Scores from labelling /m/, /n/ and /ŋ/ counted by number of participants who have given their mark.