

Multi-speaker articulatory-to-acoustic speech synthesis with deep neural networks

Bence Halpern¹, Co-author Name²

¹Netherlands Cancer Institute

²Co-author Affiliation

author@university.edu, coauthor@company.com

Abstract

For your paper to be published in the conference proceedings, you must use this document as both an instruction set and as a template into which you can type your own text. If your paper does not conform to the required format, you will be asked to fix it.

Please do not reuse your past papers as a template. To prepare your paper for submission, please always download a fresh copy of this template from the conference website and please read the format instructions in this template before you use it for your paper.

Conversion to PDF may cause problems in the resulting PDF or expose problems in your source document. Before submitting your final paper in PDF, check that the format in your paper PDF conforms to this template. Specifically, check the appearance of the title and author block, the appearance of section headings, document margins, column width, column spacing, and other features such as figure numbers, table numbers and equation number. In summary, you must proofread your final paper in PDF before submission.

The maximum number of pages is 5. The 5th page may be used exclusively for references. The references should begin on an earlier page immediately after the Acknowledgements section, and continue onto the 5th page. If no space is available on an earlier page, then the references may begin on the 5th page.

Index terms should be included as shown below.

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

Understanding how articulation works is a central question in speech research. The Birkholz model is also nice. Articulatory synthesis is a solution to this problem, but lacks sufficient speech quality. Recently, data-driven methods were discovered on small datasets.

Our method is different in slight, but important ways. However, in this paper we aim to introduce proceedings with a neural network based framework which aims to study human articulation.

Electromagnetic articulography is a tool to record positions of the articulatory trajectories.

Such a framework would be a key tool in pathological speech synthesis, because a lot of domain knowledge is known about how these articulatory trajectories effect speech.

[XY] demonstrate first the capability to use neural networks for articulatory to acoustic networks, and [X], [Y], extended their framework.

A significant difference of this framework is that we also incorporate the fundamental frequency as an estimation, because that it is independent of the articulation.

The main contributions of this paper are,

- a set of benchmarks for multi-speaker articulatory to acoustic synthesis
- estimation on how much data would be needed to construct a better model
- an example of how to construct pathological speech using this framework

2. Method

2.1. Dataset

The MNGU0 and the MOCHA-TIMIT dataset were used to obtain a total of X samples, which were partitioned to training-validation set.

2.2. Preprocessing

The delay was [TODO: ACF]

There are many techniques for, F0 preprocessing [TODO], the one that we employed was with linear interpolation.

Taking into account the different sampling rates of the two different articulographs, they were both downsampled to 200 Hz (?) and an analysis window of 5ms were used.

For the MNGU0 dataset, NaN values were of EMA occurred which were interpolated linearly.

The speech was transformed with the PyWORLD vocoder (TODO: reference the url) and compressed with the PySPTK toolkit. The resulting 40 MFCC and 1 power parameters were used to generative static and delta parameters, resulting in 82 parameters for the training. As the first step of the MFCC processing $\alpha = 0.42$ were used a pre-emphasis coefficient. The PyWORLD vocoder also provides the F_0 and BAP values for synthesis, but this was not predicted on our experiment see Section Synthesis.

Exploratory data analysis indicated that the articulatory trajectories of the datasets were on different scale and bias, so it was normalised on a per speaker basis. The output MFCCs were also normalised, but not on a per speaker basis.

2.3. Neural network

BLSTM and LSTM were done

2.4. Synthesis

A slightly different method were undertaken for synthesis. It might seem unconventional, that we don't predict BAP and F_0 , but F_0 is directly related to laryngeal function, so it is at the input side of the actual speech synthesis.

3. Experiment

4. Results and discussion

Table 1: Comparison of 10-fold CV performance of neural networks

Architecture	MCD
LSTM	6 dB
BLSTM	1000 dB

5. Conclusion

6. Acknowledgements

7. Page layout and style

Authors should observe the following rules for page layout. A highly recommended way to meet these requirements is to use a given template (Microsoft Word or LaTeX) and check details against the corresponding example PDF file. Given templates, Microsoft Word® or LaTeX, can be adapted/imported easily in other software such as LibreOffice, Apple Pages, LuaLaTeX, and XeLaTeX, but please be careful to match the layout of the provided PDF example.

7.1. Basic layout features

- Proceedings will be printed in DIN A4 format. Authors must submit their papers in DIN A4 format.
- Two columns are used except for the title section and for large figures that may need a full page width.
- Left and right margin are 20 mm each.
- Column width is 80 mm.
- Spacing between columns is 10 mm.
- Top margin is 25 mm (except for the first page which is 30 mm to the title top).
- Bottom margin is 35 mm.
- Text height (without headers and footers) is maximum 235 mm.
- Headers and footers must be left empty.
- Check indentations and spacings by comparing to this example file (in PDF).

7.1.1. Headings

Section headings are centered in boldface with the first word capitalized and the rest of the heading in lower case. Sub-headings appear like major headings, except they start at the left margin in the column. Sub-sub-headings appear like sub-headings, except they are in italics and not boldface. See the examples in this file. No more than 3 levels of headings should be used.

7.2. Text font

Times or Times Roman font is used for the main text. Font size in the main text must be 9 points, and in the References section 8 points. Other font types may be used if needed for special purposes. It is VERY IMPORTANT that while making the final PDF file, you embed all used fonts! To embed the fonts, you may use the following instructions:

1. For Windows users, the bullzip printer can convert any PDF to have embedded and subsetted fonts.
2. For Linux/Mac users, you may use
pdftops file.pdf
pstopdf -dPDFSETTINGS=/prepress file.pdf

LaTeX users: users should use Adobe Type 1 fonts such as Times or Times Roman. These are used automatically by the INTERSPEECH2018.sty style file. Authors must not use Type 3 (bitmap) fonts.

7.3. Figures

All figures must be centered on the column (or page, if the figure spans both columns). Figure captions should follow each figure and have the format given in Figure 1.

Figures should be preferably line drawings. If they contain gray levels or colors, they should be checked to print well on a high-quality non-color laser printer.

Graphics (i.e., illustrations, figures) must not use stipple fill patterns because they will not reproduce properly in Adobe PDF. Please use only SOLID FILL COLORS.

Figures which span 2 columns (i.e., occupy full page width) must be placed at the top or bottom of the page.

7.4. Tables

An example of a table is shown in Table 2. The caption text must be above the table.

Table 2: This is an example of a table

Ratio	Decibels
1/10	-20
1/1	0
2/1	≈ 6
3.16/1	10
10/1	20
100/1	40
1000/1	60

7.5. Equations

Equations should be placed on separate lines and numbered. Examples of equations are given below. Particularly,

$$x(t) = s(f_{\omega}(t)) \quad (1)$$

where $f_{\omega}(t)$ is a special warping function

$$f_{\omega}(t) = \frac{1}{2\pi j} \oint_C \frac{\nu^{-1k} d\nu}{(1 - \beta\nu^{-1})(\nu^{-1} - \beta)} \quad (2)$$

A residue theorem states that

$$\oint_C F(z) dz = 2\pi j \sum_k \text{Res}[F(z), p_k] \quad (3)$$

Applying (3) to (1), it is straightforward to see that

$$1 + 1 = \pi \quad (4)$$

Finally we have proven the secret theorem of all speech sciences. No more math is needed to show how useful the result is!

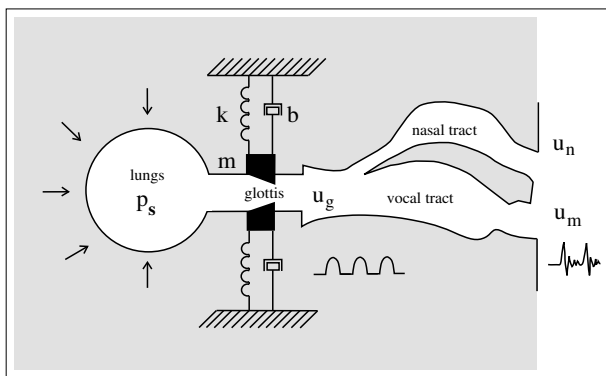


Figure 1: Schematic diagram of speech production.

If your Word document contains equations, you must not save your Word document from “.docx” to “.doc” because when doing so, Word will convert all equations to images of unacceptably low resolution.

7.6. Hyperlinks

For technical reasons, the proceedings editor will strip all active links from the papers during processing. Hyperlinks can be included in your paper, if written in full, e.g. “<http://www.foo.com/index.html>”. The link text must be all black. Please make sure that they present no problems in printing to paper.

7.7. Multimedia files

The INTERSPEECH organizing committee offers the possibility to submit multimedia files. These files are meant for audio-visual illustrations that cannot be conveyed in text, tables and graphs. Just like you would when including graphics, make sure that you have sufficient author rights to the multimedia materials that you submit for publication. The proceedings media will NOT contain readers or players, so be sure to use widely accepted file formats, such as MPEG, Windows WAVE PCM (.wav) or Windows Media Video (.wmv) using standard codecs.

Your multimedia files must be submitted in a single ZIP file for each separate paper. Within the ZIP file you can use folders and filenames to help organize the multimedia files. In the ZIP file you should include a TEXT or HTML index file which describes the purpose and significance of each multimedia file. From within the manuscript, refer to a multimedia illustration by its filename. Use short file names without blanks for clarity.

The ZIP file you submit will be included as-is in the proceedings media and will be linked to your paper in the navigation interface of the proceedings. Causal Productions (the publisher) and the conference committee will not check or change the contents of your ZIP file.

Users of the proceedings who wish to access your multimedia files will click the link to the ZIP file which will then be opened by the operating system of their computer. Access to the contents of the ZIP file will be governed entirely by the operating system of the user’s computer.

7.8. Page numbering

Final page numbers will be added later to the document electronically. *Don’t make any footers or headers!*

7.9. References

The reference format is the standard IEEE one. References should be numbered in order of appearance, for example [?], [?], [?, pp. 417–422], and [?].

7.10. Abstract

The total length of the abstract is limited to 200 words. The abstract included in your paper and the one you enter during web-based submission must be identical. Avoid non-ASCII characters or symbols as they may not display correctly in the abstract book.

7.11. Author affiliation

Please list country names as part of the affiliation for each country.

7.12. Number of authors in the author list

The maximum number of authors in the author list is twenty. If the number of contributing authors is more than twenty, they should be listed in a footnote or in acknowledgement section, as appropriate.

7.13. Submitted files

Authors are requested to submit PDF files of their manuscripts. You can use commercially available tools or for instance <http://www.pdfcreator.org/products/pdfcreator>. The PDF file should comply with the following requirements: (a) there must be no PASSWORD protection on the PDF file at all; (b) all fonts must be embedded; and (c) the file must be text searchable (do CTRL-F and try to find a common word such as the). The proceedings editors (Causal Productions) will contact authors of non-complying files to obtain a replacement. In order not to endanger the preparation of the proceedings, papers for which a replacement is not provided timely will be withdrawn.

8. Discussion

This is the discussion. This is the discussion. This is the discussion. Is there any discussion?

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras consequat mollis odio, nec venenatis enim auctor sed. Integer tincidunt fringilla lectus eget condimentum. In eget sapien id eros dapibus interdum vel ac quam. Aenean vitae rutrum erat. Aenean et risus pharetra, lacinia augue ut, fermentum ante. Integer dui arcu, interdum at ornare a, faucibus quis est. Mauris quis quam felis. Etiam pulvinar massa et turpis lacinia, eu posuere mi iaculis. Fusce at velit quis leo dignissim porttitor.

Fusce ut nunc eu sapien venenatis finibus a vel ligula. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Ut quam eros, volutpat at gravida consectetur, rutrum ut leo. Aenean cursus euismod feugiat. Cras hendrerit, ligula eu feugiat malesuada, neque turpis auctor lacus, sit amet accumsan neque orci a quam. Mauris suscipit ultrices mattis. Nulla at interdum metus, id pharetra diam. Curabitur at vestibulum sem, sed elementum massa. Donec iaculis et arcu ut rutrum. Fusce gravida, mauris porta volutpat eleifend, enim mauris eleifend orci, eu ultrices leo purus vitae metus. In pretium dolor ut magna dictum, at imperdiet lectus porta.

Quisque mollis lectus id risus pretium mattis. Morbi scelerisque posuere est, id efficitur urna luctus non. Praesent quam lacus, facilisis id ante eu, vehicula maximus ex. Nullam

mollis in arcu vitae efficitur. Aliquam molestie eleifend ante, in pretium velit ultrices ac. Etiam laoreet nec sem non pulvinar. Integer ligula felis, interdum non lacus id, malesuada imperdiet turpis.

Aenean sit amet volutpat nisi. Aliquam eu erat quis tortor ultrices laoreet. Vivamus fermentum semper metus, non faucibus libero euismod vitae. Sed efficitur porta congue. Aenean in faucibus nisi. Donec suscipit augue vitae orci consequat, sit amet aliquet felis varius. Duis efficitur lacinia dolor sit amet lobortis. Curabitur erat sapien, molestie nec nisi eu, dignissim accumsan ipsum. Fusce id nibh nec risus dictum posuere in ac magna. Donec malesuada massa sed erat lacinia cursus. Suspendisse ornare augue nec volutpat consequat.

Vestibulum et vulputate nisi, a malesuada mi. Nam pellentesque arcu sapien, at placerat odio imperdiet ut. Curabitur nec venenatis tellus, vel aliquet nisi. Curabitur vel ligula sit amet metus auctor pretium. Nullam nulla mi, blandit a mattis id, vulputate sit amet enim. Proin mollis fringilla dictum. Proin lacinia orci purus.

Curabitur porttitor bibendum dolor, nec consectetur sapien pulvinar id. Donec eleifend, est vel dignissim pretium, tortor augue euismod nunc, id fermentum erat felis ac neque. Morbi id lectus ultricies, rutrum justo eu, sollicitudin risus. Suspendisse lobortis efficitur nisi sit amet pellentesque. Ut eget augue at mi aliquet mattis. Proin et feugiat erat, sit amet sodales eros. Integer sed elit quis est mattis ullamcorper. Pellentesque lectus nisi, vulputate a imperdiet tincidunt, auctor nec orci. Pellentesque sagittis nisl orci, vitae placerat massa lacinia nec. Sed egestas magna sed augue sollicitudin luctus. Praesent interdum bibendum tortor, eu porta purus. Aliquam convallis velit id mi fermentum, sed ornare eros cursus.

Quisque congue leo a fringilla pharetra. Phasellus sed tempor est, sed auctor purus. Morbi vel lacus ullamcorper, auctor mauris id, pulvinar lorem. Suspendisse potenti. Nam porta, purus non eleifend bibendum, erat metus pellentesque elit, non luctus nibh nunc ornare nisl. Sed rutrum lacinia nisi ac suscipit. Curabitur non blandit augue. Integer viverra, ipsum vel molestie euismod, sem quam tempus massa, eget efficitur ante turpis non metus. Quisque efficitur posuere velit in iaculis. Cras imperdiet varius urna vitae vestibulum. Donec accumsan eget nisi sed pellentesque. Vestibulum id quam ut urna volutpat ullamcorper gravida sit amet libero. Aliquam bibendum, ligula vitae porta malesuada, arcu diam congue erat, a pharetra diam sem vulputate tortor. Etiam luctus iaculis leo cursus tristique.

Mauris mattis sem dolor, sit amet ullamcorper arcu tincidunt ac. Vestibulum at blandit tortor. Quisque bibendum congue leo, vitae eleifend massa. Vestibulum vitae odio elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut sagittis quam vel felis ornare, in gravida felis tempor. Donec molestie dui quis leo venenatis blandit. Nunc sit amet finibus metus. Cras ut lectus ex.

Suspendisse commodo libero vel leo tincidunt, a tempus mauris porta. Integer varius eros ac sapien lacinia vehicula. Donec porttitor, lacus faucibus rhoncus venenatis, neque quam imperdiet nunc, id consectetur metus purus quis sapien. Phasellus interdum nulla vel euismod posuere. Vestibulum finibus magna vel finibus mollis. Curabitur mollis turpis tortor, hendrerit vulputate justo egestas quis. Nam dignissim luctus leo non elementum. Phasellus a metus at leo malesuada bibendum. Mauris quis eleifend magna, nec vehicula ex. Donec venenatis urna fermentum commodo vehicula. Ut mattis scelerisque aliquam. Vivamus pulvinar erat metus, id tempus mi vulputate quis.

Fusce lobortis a urna eget blandit. Vivamus in eleifend

neque, at sollicitudin lectus. Quisque faucibus egestas lorem, in commodo diam maximus eu. Morbi finibus ante ac felis porttitor euismod. Donec lobortis aliquam ipsum sit amet luctus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam rutrum neque sapien, eget luctus turpis iaculis pulvinar. Duis quis pulvinar nunc, nec bibendum ligula. Phasellus suscipit sagittis lacus molestie laoreet. Pellentesque lacus diam, tincidunt a aliquam vitae, aliquet non justo.

Etiam lectus lacus, commodo eget consectetur eget, auctor vitae leo. Praesent vitae erat in diam blandit semper vitae et eros. Maecenas auctor pharetra nibh eget egestas. Donec accumsan ut risus eget rhoncus. Nam placerat, erat sit amet gravida mollis, purus arcu accumsan diam, tempus pharetra risus mi ac sapien. Ut et tortor porta, pulvinar elit vitae, tempor mi. Nam interdum, nisl non pharetra molestie, turpis neque commodo ligula, sit amet pretium nisl nibh quis ante. Quisque et ex eget velit lobortis suscipit. Integer aliquam finibus molestie. Sed pellentesque neque eu turpis aliquet, mattis ornare enim finibus. In hac habitasse platea dictumst.

Integer congue quis justo a posuere. Quisque porta, ante et dignissim suscipit, arcu mauris ultrices libero, nec sollicitudin purus lacus a enim. Aliquam feugiat eget lacus molestie sodales. Duis blandit placerat nunc, et venenatis turpis dictum vel. Nulla facilisi. Nam ullamcorper, tellus eu posuere mattis, arcu lacus dictum nulla, vel mattis nisi sem posuere tellus. Etiam quis eros condimentum lectus lobortis eleifend. In ex lacus, sodales scelerisque egestas ac, aliquam nec purus. Nunc sit amet magna non libero ullamcorper dictum. Phasellus porta faucibus tempus. Praesent blandit tortor sed tellus ornare consectetur. Sed sed nisi id neque porta varius eu eu velit. Curabitur varius convallis justo id facilisis. Mauris auctor velit nec aliquam cursus.

Integer suscipit scelerisque leo sed faucibus. Ut commodo nulla luctus diam posuere egestas. Integer ut augue ac velit ullamcorper tempus. Pellentesque in mi rhoncus, sodales sem quis, commodo sem. Aenean dapibus euismod diam id rhoncus. Nullam vehicula placerat eros consectetur luctus. Aliquam auctor ipsum vitae egestas imperdiet. Ut nulla lacus, imperdiet quis urna vel, ornare imperdiet tortor. Mauris nec diam ac nunc laoreet volutpat at id turpis. Nulla eu neque a risus feugiat iaculis ac vel risus. Ut tempus elementum lorem eget porta. Nullam et ullamcorper urna. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Phasellus eget dui vitae nulla hendrerit ultrices quis rutrum leo.

Proin consectetur lacus sit amet eleifend varius. Etiam eu blandit risus. Curabitur pellentesque urna sed dolor congue mattis. Vestibulum ut velit posuere, feugiat leo a, dignissim massa. Proin eu nulla risus. Fusce luctus bibendum est, sit amet venenatis nisi finibus non. Donec ultricies ornare nunc at lobortis. Nam a auctor metus. Vestibulum pretium condimentum turpis ac mattis. Curabitur semper sagittis rhoncus. Duis molestie facilisis mattis. Sed pharetra lorem id tortor efficitur, sed maximus leo posuere. Quisque suscipit molestie convallis. Duis imperdiet placerat congue. Morbi placerat, velit ut tempor porta, ex nisi imperdiet purus, non feugiat ex velit nec nulla.

Phasellus mattis at erat eget lobortis. Vivamus sodales odio non erat luctus faucibus. Curabitur aliquam luctus nulla quis consectetur. Fusce vulputate finibus vulputate. Cras at condimentum massa. Duis vestibulum ipsum ac tortor lobortis fringilla. Cras non neque at nunc pellentesque mollis. Sed quis erat mauris. Ut dapibus sem lectus, quis imperdiet diam bibendum et. Maecenas quis venenatis ante. Nunc blandit a risus sed scelerisque. Praesent cursus est sit amet nisi tempus, quis placerat libero rutrum. Phasellus lacinia nisi quis consequat mollis.

Phasellus sagittis aliquam lacus.

Sed dolor quam, posuere nec nunc eget, feugiat lobortis ligula. Fusce lacinia fermentum dolor, luctus dapibus ex venenatis feugiat. In hac habitasse platea dictumst. Nullam vitae ligula dignissim, interdum turpis quis, tincidunt metus. Nullam in nisl vitae mauris egestas porta. Nam fringilla aliquet sapien, non dapibus nunc sollicitudin id. Nunc hendrerit felis et vehicula consequat. Praesent varius libero id volutpat iaculis. Aliquam vel dui imperdiet, pharetra augue sed, iaculis nulla. Quisque mollis orci nec odio eleifend, eget laoreet nunc feugiat. Cras varius tortor a fringilla gravida.

Sed posuere erat eu dolor consequat euismod. Donec imperdiet, tellus nec convallis commodo, lorem sem lobortis purus, a lacinia massa ipsum vitae nisl. Vivamus auctor tellus in urna iaculis luctus. In dui nibh, posuere a erat a, lobortis finibus nulla. Sed vel suscipit nisi. Nunc eget nibh risus. Sed posuere tempus eleifend. Nullam ac lacinia ligula, ut blandit erat. In at est sed turpis consequat rutrum. Nunc eget lectus venenatis, convallis ligula non, mollis orci. Aliquam sit amet ligula turpis. Sed finibus laoreet elit nec molestie.

Mauris in nisi et neque euismod aliquam ut eget felis. Sed eget dictum tellus, finibus rutrum nibh. Fusce placerat augue a faucibus semper. Nam sed nisl ligula. Vivamus ante augue, faucibus at risus ut, hendrerit viverra risus. Etiam justo eros, dignissim a nunc sed, porta mollis erat. Fusce in sem accumsan, laoreet nibh porta, porttitor orci. Donec venenatis vehicula ante eget dictum.

Aliquam fermentum a metus pellentesque cursus. Vivamus eleifend ultricies tellus, vel scelerisque diam. Sed hendrerit est at elit suscipit placerat. Praesent nec pretium erat. Quisque blandit nunc in felis bibendum consequat. Aliquam quis orci consectetur nulla luctus ullamcorper. Suspendisse finibus luctus erat a dapibus.

9. Conclusions

Authors must proofread their PDF file prior to submission to ensure it is correct. Authors should not rely on proofreading the Word file. Please proofread the PDF file before it is submitted.

10. Acknowledgements

The ISCA Board would like to thank the organizing committees of the past INTERSPEECH conferences for their help and for kindly providing the template files.

Note to authors: Authors should not use logos in acknowledgement section; rather authors should acknowledge corporations by naming them only.