JOURNAL OF DATA-DRIVEN ENGINEERING SYSTEMS

Adaptive Noise Cancellation for Speech Recorded During Magnetic Resonance Imaging: Formant Extraction and Analysis

Jack Laub[1*], Thomass Rofsky[2], Sara Weinreb[3], Dave Lee[3]

[1] Department of Mathematics and Systems Analysis, Western Norway University of Applied Sciences, Finland
[2] Department of Signal Processing and Acoustics, Western Norway University of Applied Sciences, Finland
[3] Department of Mathematics and Systems Analysis, Western Norway University of Applied Sciences, Finland

[*] Jacklaub@gmail.com

(Received: 15th June 2021; Accepted: 6th September 2021; Published on-line: 8th September 2021)

Abstract

Speech recordings obtained during Magnetic Resonance Imaging (MRI) of the upper airways often suffer from acoustic noise generated by the MRI scanner. This paper focuses on the post-processing of such speech samples using adaptive comb filtering to achieve accurate formant extraction. Two types of speech materials were used to validate the proposed algorithm: prolonged vowel productions recorded during MRI and comparison data recorded in an anechoic chamber. Spectral envelopes and vowel formants were computed from the post-processed speech and the comparison data. Additionally, numerical acoustic models and 3D printed vocal tract physical models were used for further analysis. The results reveal a significant frequency-dependent discrepancy between the vowel formant data obtained from recordings during MRI and the comparison data. This discrepancy is attributed to the acoustical changes caused by the surfaces of the MRI head coil, leading to "exterior formants" at frequencies around 1 kHz and 2 kHz. The observed discrepancy is too substantial to be disregarded when using MRI recordings for parameter estimation or validating numerical speech models based on MR images. However, the influence of test subject adaptation to noise and the effects of constrained space acoustics during an MRI examination cannot be ruled out.

Keywords: speech recording, magnetic resonance imaging, noise reduction, adaptive comb filtering, formant extraction, vocal tract, acoustic artifacts.

I. Introduction

Modern medical imaging technologies, including Ultrasonography (USG), X-ray Computer Tomography (CT), and Magnetic Resonance Imaging (MRI), have revolutionized studies of speech and articulation. While each technology has its own applicability and image quality characteristics, MRI remains an attractive approach for large-scale articulation studies due to its potential advantages. However, there are several restrictions and challenges associated with conducting speech studies during MRI scans, as discussed in previous works [1, 2]. To obtain paired data that can facilitate the development and validation of computational speech models, it is desirable to simultaneously record speech during MRI experiments. Unfortunately, the speech signal recorded during MRI is often contaminated by artifacts primarily caused by the high acoustic noise levels within the MRI scanner. Additional challenges arise due to the non-flat frequency response of the MRI-proof audio measurement system and the constrained space acoustics within the MRI head and neck coils.

In the field of signal processing, noise cancellation techniques are commonly used to enhance speech quality. Two main classes of noise cancellation methods are adaptive noise cancellation and blind source separation techniques, such as FastICA [4]. This article focuses on introducing, analyzing, and validating an adaptive noise cancellation algorithm for speech recorded during MRI. Compared to blind source separation techniques, adaptive noise cancellation offers better tractability and can be implemented in the time domain, frequency domain, or a combination of both. The algorithm discussed in this article is designed based on insights gained from a previous algorithm introduced in [2, Section 4]. Other approaches for addressing MRI noise are also discussed in [5, 6, 7, 8], which will be explored in detail towards the end of this article.

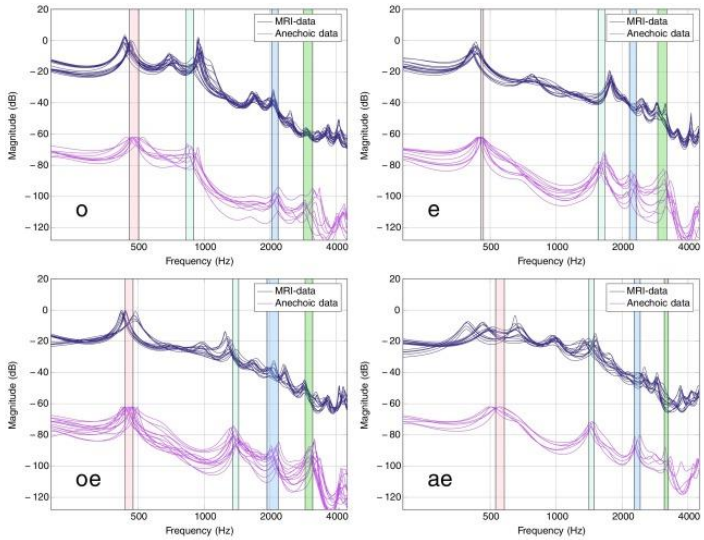

Fig 8: Spectral envelopes of all vowel

REFERENCES

[1] Aalto D, Aaltonen O, Happonen RP, Jääsaari P, Kivelä A, Kuortti J, Luukinen JM, Malinen J, Murtola T, Parkkola R, Saunavaara J. Large scale data acquisition of simultaneous MRI and speech. Applied Acoustics. 2014 Sep 1;83:64-75.

[2] Birkholz P, Kürbis S, Stone S, Häsner P, Blandin R, Fleischer M. Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties. Scientific data. 2020 Aug 5;7(1):255.

[3] Arnela M, Dabbaghchian S, Blandin R, Guasch O, Engwall O, Van Hirtum A, Pelorson X. Influence of vocal tract geometry simplifications on the numerical simulation of vowel sounds. The Journal of the Acoustical Society of America. 2016 Sep 15;140(3):1707-18.

[4] S. Izadi, K. Jabari, M. Izadi, B. Khadem Hamedani, and A. Ghaffari, "Identification and Diagnosis of Dynamic and Static Misalignment in Induction Motor Using Unscented Kalman Filter," in 2021 13th Iranian Conference on Electrical Engineering and Computer Science (ICEESC), 2021.

[5] Alku P, Murtola T, Malinen J, Kuortti J, Story B, Airaksinen M, Salmi M, Vilkman E, Geneid A. OPENGLOT–An open environment for the evaluation of glottal inverse filtering. Speech Communication. 2019 Feb 1;107:38-47.

[6] Brandner M, Blandin R, Frank M, Sontacchi A. A pilot study on the influence of mouth configuration and torso on singing voice directivity. The Journal of the Acoustical Society of America. 2020 Sep 4;148(3):1169-80.

[7] Arnela M, Dabbaghchian S, Guasch O, Engwall O. MRI-based vocal tract representations for the three-dimensional finite element synthesis of diphthongs. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019 Sep 19;27(12):2173-82.

[8] Guasch O, Arnela M, Codina R, Espinoza H. A stabilized finite element method for the mixed wave equation in an ALE framework with application to diphthong production. Acta Acustica united with Acustica. 2016 Jan 1;102(1):94-106.

[9] Schickhofer L, Malinen J, Mihaescu M. Compressible flow simulations of voiced speech using rigid vocal tract geometries acquired by MRI. The Journal of the Acoustical Society of America. 2019 Apr 16;145(4):2049-61.

[10] Arnela M, Guasch O. Finite element synthesis of diphthongs using tuned two-dimensional vocal tracts. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2017 Aug 2;25(10):2013-23.

[11] Freixes M, Arnela M, Socoró JC, Alías F, Guasch O. Glottal source contribution to higher order modes in the finite element synthesis of vowels. Applied Sciences. 2019 Oct 25;9(21):4535.

[12] Schickhofer L, Mihaescu M. Analysis of the aerodynamic sound of speech through static vocal tract models of various glottal shapes. Journal of biomechanics. 2020 Jan 23;99:109484.