Physica Medica 46 (2018) 96–103

Contents lists available at ScienceDirect

Physica Medica

European Journal of Medical Physics

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Real-time speech MRI: Commercial Cartesian and non-Cartesian sequences at 3T and feasibility of offline TGV reconstruction to visualise velopharyngeal motion

Andreia C. Freitas[a,b], Matthieu Ruthven[b], Redha Boubertakh[a,b], Marc E. Miquel[a,b,*]

[a] *William Harvey Research Institute, Queen Mary University of London, Charterhouse Square, EC1M 6BQ London, United Kingdom*
[b] *Clinical Physics, Barts Health NHS Trust, London, United Kingdom*

ARTICLE INFO

Keywords:
Velopharyngeal closure
Real-time MRI
Speech
Non-Cartesian

ABSTRACT

Purpose: This study aims to improve clinical reliability of real-time Magnetic Resonance Imaging (rt-MRI) in the visualisation of velopharyngeal motion during speech.
Methods: Seven subjects were imaged at 3T during natural phonation. Speech rt-MRI methodologies were investigated with (i) a comparison of commercial Cartesian and non-Cartesian (radial and spiral) rt-MRI sequences and (ii) investigation of further improvement with accelerated radial acquisition and offline reconstruction methodology.
Results: Cartesian and non-Cartesian protocols were implemented with temporal resolutions between 10 frames per second (fps) and 27 fps and voxel sizes between $1.5 \times 1.5 \times 10\,mm^3$ and $2.7 \times 2.7 \times 10\,mm^3$. Commercial spiral acquisitions provided superior contrast-to-noise ratio (CNR) than otherwise equivalent Cartesian and radial. Spirals at 22 fps allowed for improved spatial resolution ($1.9 \times 1.9\,mm^2$) when compared to similar Cartesian protocols (20 fps), limited to a lower spatial resolution ($2.7 \times 2.7\,mm^2$). Cartesian protocols were on average scored higher than spiral protocols in visual quality. However, some variability was found on choice of recommended imaging protocol between subjects. Accelerated radial data reconstructed offline with a Total Generalized Variation (TGV) scheme showed improved visual sharpness of velum motion.
Discussion/Conclusion: Adequate visualisation of velopharyngeal motion with commercial rt-MRI at 3T was possible. Both Cartesian and spiral protocols demonstrated adequate temporal depiction and overall image quality. However, choice of optimal imaging protocol at 3T was more subject-dependent than in previously published 1.5T data and additional care should be taken when selecting an adequate protocol. Offline TGV reconstruction of radial data has shown potential to improve temporal sharpness.

1. Introduction

The production of human speech relies on the complex and timely co-articulation of several vocal structures. One of the processes during speech production, and the focus of this study, is the closure of the velopharyngeal port. Velopharyngeal closure is characterised by the posterior-elevation of the soft palate (or velum) and the approximation of the lateral pharyngeal walls. However, patients with a range of different speech disabilities frequently present velopharyngeal insufficiency (VPI), i.e. the incomplete closure of the velopharyngeal port. As a result, air escapes through the nasal cavity and patients with VPI commonly present hypernasal speech [1]. Clinical assessment of these patients by speech and language therapists involves X-ray videofluoroscopy and/or nasendoscopy [1]. However, these techniques present some disadvantages, such as exposure to ionising radiation, limited soft tissue contrast and invasiveness [1]. Limitations to these techniques have strongly supported the use of Magnetic Resonance Imaging (MRI) to dynamically image the upper vocal tract during speech, as summarised in a review publication of the field [2].

Due to the continuous motion of the vocal tract structures during speech, and its importance in clinical speech assessment, adequate temporal resolution is a key issue. Standard assessment in the UK uses X-ray videofluoroscopy, usually acquired at 15 frames per second (fps). However, videofluoroscopy examinations are intrinsically limited in frame rate, total duration (recommended up to 2 min) and number of scan repetitions, in order to minimise patient exposure to ionising radiation [1].

Previous studies on assessing velopharyngeal closure with real-time

* Corresponding author at: Clinical Physics, St Bartholomew's Hospital, Barts Health NHS Trust, London, EC1A 7BE, UK.
E-mail address: m.e.miquel@qmul.ac.uk (M.E. Miquel).

https://doi.org/10.1016/j.ejmp.2018.01.014
Received 1 June 2017; Received in revised form 20 January 2018; Accepted 23 January 2018
1120-1797/

MRI (rt-MRI) have used a wide range of temporal resolutions, with dynamic frames acquired every 50–200 ms [3–5]. However, because of the risk of inadequate representation of velar motion due to insufficient frame rate, a recommendation paper in the field recommends a target zone of 100 ms (10 fps) [6]. In addition, the results of a recent rt-MRI speech study investigating the frame rate required to capture velopharyngeal closures in several speech tasks suggests that a median temporal resolution of 5–7.5 fps and 7.5–10 fps is required for natural and faster speech rates [7]. Contrary to temporal resolution, not much work has been shown on an adequate in-plane spatial resolution to be considered. However, careful consideration should still be taken given due to the small dimensions of some of the organs at interest. Previous studies have reported pixel areas between 2.25 mm^2 (1.5 × 1.5 mm^2) and 5.3 mm^2 (3.3 × 1.6 mm^2) in a variety of cohorts (e.g. adults and children) [3–5,8,9].

Several studies have successfully imaged the upper vocal tract region with standard commercially available sequences and on-the-fly reconstruction schemes. Scott et al. [4] investigated velopharyngeal closure up to 50 ms (20 fps) using Cartesian Steady State Free Precession (SSFP) sequences and parallel imaging (SENSitivity Encoding – SENSE [10]). Martins et al. [8] investigated velum motion at 14 fps (3.3 × 1.6 × 8 mm^3) with Cartesian and parallel imaging Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA [11]). Turbo Spin Echo (TSE) sequences with partial Fourier imaging were used by Beer et al. [3] to achieve about 5–6 fps (1.5 × 3.1 × 6 mm^3). In this case, the performance of rt-MRI to visualise velopharyngeal closure was compared to the use of standard X-ray videofluoroscopy in a cohort of VPI patients. Imaging at higher field strengths may provide improvements in signal-to-noise (SNR), however, image quality might also be diminished due to artefacts caused by stronger off-resonance effects at the air-tissue boundary. A study by Kulinna-Cosentini et al. [12] compared the performance of TrueFISP and HASTE (single shot turbo spin echo) sequences at 3T in a cleft palate cohort, however, at suboptimal frame rates (~2.75 fps).

In addition, non-Cartesian trajectories have shown great potential to further improve spatial-temporal resolution. Narayanan et al. [13] imaged the upper vocal tract at a native frame rate of 9 fps (1.8 × 1.8 × 5 mm^3) reconstructed to 24 fps with a view sharing (sliding window) scheme. Also with a spiral acquisition, improvements in native frame rate [5] (21.4 fps reconstructed to 30 fps with view sharing) and mitigation of off-resonance artefacts [14] were later suggested. Freitas et al. [15] investigated velopharyngeal motion with commercial non-Cartesian (radial and spiral) spoiled gradient-echo sequences and on-the-fly view sharing reconstruction in comparison to the performance of Cartesian SENSE at 1.5T. An improved spiral protocol at 22 fps (1.9 × 1.9 mm^2) with superior dynamic CNR and image quality was suggested for that field strength.

Further improvements have also been reported with the use of offline iterative reconstruction methods. Burdumy et al. [16] recently performed morphometric measurements of the vocal tract articulators using an accelerated golden-angle radial sequence and offline Conjugate Gradient SENSE (CG-SENSE) reconstruction pipeline (25 fps and 1.8 × 1.8 mm^2). Similarly, Lingala et al. [17] presented a multi-shot golden-angle spiral sampling and offline sparse SENSE constrained reconstruction to further improve spatial-temporal resolution (12 ms per single-slice and 36 ms for three-slice imaging, at 2.4 × 2.4 mm^2). Niebergall et al. [9] additionally performed dynamic imaging of speech at 30 fps with a radial sampling and regularized inverse reconstruction method. These methods provide improved spatial-temporal resolutions, however, they are inherently more computationally complex and require considerable offline processing time. Although valuable in speech research studies, direct translation to clinical assessment is still challenging, in particular if a simultaneous assessment carried out by a speech and language therapist during scanning is intended. Consequently, many groups have preferred some sort of on-the-fly reconstruction (for real-time scan planning and adjustment) with posterior reconstruction and/or processing offline.

Dynamic imaging of the vocal tract with rt-MRI is still an open field of research and much variability is seen in preferred sequences, reconstruction pipelines, choice of parameters, etc. [6]. In the present study, we intend to improve clinical reliability of velopharyngeal closure visualisation with rt-MRI by providing: (i) a comparison of image quality and temporal resolution trade-off of commercial Cartesian and non-Cartesian sequences at 3T with on-the-fly reconstruction and (ii) investigating further improvements with posterior offline processing and reconstruction of accelerated non-Cartesian data. A comparison of commercial and widely available protocols may provide a starting point for researchers entering the field and aid in direct translation to the use of rt-MRI in clinical assessment, not currently standard.

2. Methods

2.1. Subjects and speech task

Seven healthy adult subjects (2 females and 5 males, median age 42 years) were recruited and imaged for this study. All volunteers gave informed written consent, in accordance with ethics committee requirements. None of the selected subjects had any known speech, hearing or language disorder. A speech task consisting of counting from 1 to 10, non-sense vocalisation (/za-na-za/, /zu-nu-zu/ and /zi-ni-zi/) and sustained phonation of the sounds /a/ and /i/ was considered [15]. Each speech sample was only repeated once per dynamic acquisition. Participants were instructed to perform a couple of nasal breaths at the beginning of each acquisition (allowing complete lowering of the velum) and to perform the speech task at a natural speech rate.

2.2. Comparison of commercially available protocols

Subjects were imaged in the supine position using a 3T Tx Philips Achieva (Philips Healthcare, Best, the Netherlands, software release 3.2) MRI scanner and a 16-channel neurovascular coil by the same manufacturer. Real-time 2D mid-sagittal images of the head and neck, centred on the velum region, were acquired.

A first set of experiments was set to compare real-time Cartesian and non-Cartesian protocols when assessing velopharyngeal closure during speech. Preliminary studies using a phantom and 2 subjects of the original cohort were first undertaken in order to identify and optimise non-Cartesian acquisition protocols. Although further improvement in spatial resolution could have been investigated, the main goal at this stage was to compare imaging protocols in regards to image quality and temporal resolution compromise under otherwise similar conditions. Therefore, non-Cartesian sequences were matched in spatial-temporal resolution to previously published Cartesian protocols [4]. Two further non-Cartesian protocols (at 41 and 38 ms) were also implemented to investigate further improvements. Acquisition parameters can be found in Table 1. Cartesian data was acquired using SSFP sequences (described here [4]). As previously established at 1.5T [15], SSFP non-Cartesian acquisitions exhibited poor image quality and consequently FLASH like sequences, commonly used in speech MRI [2], were chosen for non-Cartesian protocols. A 10 mm thick single slice was used in every scan.

On-the-fly data reconstruction was accomplished using vendor available software (integrated in the 3T Philips Achieva scanner system) to allow for immediate planning and guidance during acquisition. Cartesian data were combined using SENSE while non-Cartesian data was reconstructed with a standard gridding and view sharing method, also available in the scanner system. No further data post-processing was performed at this stage; DICOM image data were exported and analysed.