Physica Medica 30 (2014) 604–618

Contents lists available at ScienceDirect

Physica Medica

journal homepage: http://www.physicamedica.com

Review paper

Speech MRI: Morphology and function

Andrew D. Scott [a,b], Marzena Wylezinska [a,c], Malcolm J. Birch [a], Marc E. Miquel [a,c,*]

[a] Clinical Physics, Barts Health NHS Trust, London EC1A 7BE, United Kingdom
[b] NIHR Cardiovascular Biomedical Research Unit, The Royal Brompton Hospital, Sydney Street, London SW3 6NP, United Kingdom
[c] Barts and The London NIHR CVBRU, London Chest Hospital, London E2 9JX, United Kingdom

ARTICLE INFO

Article history:
Received 3 March 2014
Received in revised form
24 April 2014
Accepted 1 May 2014
Available online 28 May 2014

Keywords:
Speech imaging
MRI dynamic imaging
Real-time imaging
Vocal tract imaging and modelling
Cleft palate
Vocal tract morphology
Speech production

ABSTRACT

Magnetic Resonance Imaging (MRI) plays an increasing role in the study of speech. This article reviews the MRI literature of anatomical imaging, imaging for acoustic modelling and dynamic imaging. It describes existing imaging techniques attempting to meet the challenges of imaging the upper airway during speech and examines the remaining hurdles and future research directions.

Introduction

Basics of speech production

The production of human speech is a complex process, see Refs. [1–4] for example, involving numerous organs, namely: the lungs, diaphragm and chest wall; the larynx, pharynx and vocal folds (or cords); the tongue, lips and soft palate (or velum); and the teeth, jaw and nasal cavity. The lungs, driven by the diaphragm and/or chest wall provide the airflow which travels through the lower respiratory tract (the bronchioles, the two bronchial tubes and the trachea). Air enters the upper respiratory tract (which will be the focus of this review) through the larynx which contains the vocal cords, a schematic of the upper airway in mid-sagittal view is represented on Fig. 1. Air is forced through a narrow gap between the vocal folds which vibrate, producing a fundamental frequency plus harmonics. By manipulation of the tension, length and separation of the vocal cords and control of the airflow between them, the fundamental frequency, volume, and therefore the intonation, of speech can be controlled. The remainder of the upper respiratory tract forms a series of connected resonant cavities which can be modified in size and shape using the pharynx, velum, uvula, jaw, tongue and lips. These manipulators modify the formant frequencies of speech. For consonant sounds, speech is further complicated by articulation, which is the partial or full obstruction of the vocal tract by a pair of articulators – the tongue tip and upper teeth, tongue body and hard palate, the lips or the velum and tongue dorsum, for example (Fig. 2).

Clinical assessment of speech

A number of acquired and inherited diseases can affect the function of the speech organs, including cancer [5,6], clefts of the lips and/or palate [4], laryngitis, vocal cord polyps [7], nodules or cysts [8], neurological conditions [9] and some endocrine disorders [10]. Standard diagnostic tools for speech therapists assessing speech disorders include aural assessment, acoustic analysis, oral and nasal airflow measurements [11] and external measurement of the impedance across the vocal folds to measure their movement (electroglottography) [12]. Imaging techniques may also be used, most commonly endoscopy. Stroboscopy is a variant of endoscopy which is used for analysis of vocal fold movement in slow motion [13]. X-Ray fluoroscopy is also used to assess the movement of the tongue [14], soft palate and pharyngeal wall movements [15,16]. However, endoscopy is invasive (although minimally) potentially causing discomfort and abnormal speech. It is also limited to

* Corresponding author. Clinical Physics, St Bartholomew's Hospital, Barts Health NHS Trust, London EC1A 7BE, United Kingdom. Tel.: +44 203 465 6771; fax: +44 203 465 5785.
E-mail address: m.e.miquel@qmul.ac.uk (M.E. Miquel).

http://dx.doi.org/10.1016/j.ejmp.2014.05.001
1120-1797/

produce en-face images of exterior surfaces within the vocal tract. In contrast, X-ray fluoroscopy is non-invasive and operates at both high temporal (~30 frames per second (fps)) and high in-plane spatial resolution (<0.5 × 0.5 mm^2 [17]), but is limited to produce projections of the anatomy. The soft-tissue contrast of X-ray fluoroscopy is relatively poor, but the bony structures are clearly seen and the outline of both the soft palate and posterior pharyngeal wall can be identified. Typically, the lateral view is used, but other orientations such as the townes and basal views [15,18,19] have been used to add additional information. Increased contrast [20] is available by coating the vocal tract with barium contrast at the expense of increased patient discomfort [21]. X-Ray imaging also results in an ionising radiation dose to both the patient and operator [22]. Estimates of the average patient effective dose in dysphagia studies, which target similar anatomy, vary between 0.2 mSv [23], 0.4 mSv [24] and 0.85 mSv [22], but are dependent on the exposure time. Speech assessments may occur regularly, resulting in repeated radiation exposures and studies are typically longer than dysphasia protocols in order to obtain a varied speech sample.

Alternative imaging techniques: CT and ultrasound

Other imaging techniques available for imaging the speech articulators include ultrasound and X-ray computed tomography (CT). Ultrasound is relatively low cost, widely available, rapid and free from ionising radiation [25]. Relatively narrow (~2 mm) sections of the mid part of the tongue can be imaged in coronal or sagittal planes, from underneath the chin, at temporal resolutions of around 30 fps and sub-millimetre spatial resolution. However, while the palate and velum may be observed in some lingual positions, air interfaces, and bone are highly reflective and structures beyond these interfaces are not visualised, somewhat limiting the technique. Furthermore, the need for the transducer to be in direct contact with the skin may interfere with normal speech.

CT is able to obtain high spatial resolution (<1 × 1 × 1 mm^3) volumetric images with delineation of soft tissue and bone, which may be reformatted in any plane. However, radiation doses are typically greater than those from planar X-ray studies. Furthermore, temporal resolution is currently limited to around 165 ms at 3 gantry rotations per second or around 85 ms using state of the art dual source CT systems [26,27] which is insufficient for detailed analysis of tongue, lip and velar motion.

MRI: imaging anatomy and function?

MR imaging provides tomographic images with excellent soft tissue contrast in any plane without the use of ionising radiation and scanners are now commonplace. While MR was previously considered as a “slow” imaging modality, modern techniques, largely developed as a result of the desire to capture or freeze the motion of the heart, can result in temporal resolutions far exceeding those available with CT and even ultrasound [28]. As it is also able to provide images of both the anatomy and function of the vocal tract, MRI as the potential to become the modality of choice in speech imaging. This has resulted in a large body of literature describing the acquisition of MR images in the vocal tract and the potential diagnostic and modelling applications.

Despite the growing body of literature, outside technical MR journals many studies fail to adequately or accurately describe the MR acquisition parameters. Studies often neglect to fully describe important parameters such as resolution [29–34], flip angle [33,35–41], echo train length [33,42–45], or sequence type [30,46,47]. Errors in the description of the methods include echo times (TE) longer than repetition times (TR) [48], descriptions of

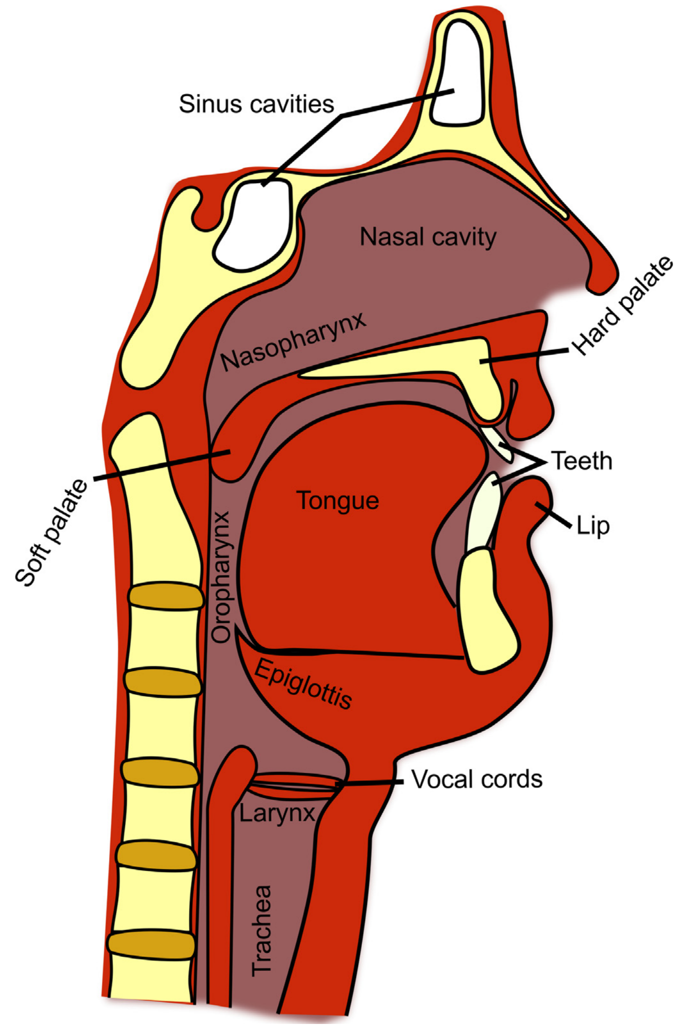

Figure 1. Anatomy of the upper airways, mid-sagittal view.

multi-slice 2D sequences as 3D acquisitions [46] and a steady state turbo spin echo sequence (instead of single shot) [49].

While existing work has considered the application of MRI in a variety of situations, challenges and limitations still exist and further work is required before the modality is used routinely in the vocal tract. For example, large changes in magnetic susceptibility are present at tissue–air interfaces which result in sequence dependent artefacts; signal to noise ratio (SNR) must be traded for temporal and/or spatial resolution; many of the reconstruction techniques for rapid acquisitions are computationally challenging; and imaging is almost universally performed with the subject in the supine position.

No previous review of this large and expanding field has previously been published. In this review we will attempt to provide a rational approach to the body of literature divided into imaging anatomy, imaging for acoustic modelling and dynamic imaging. We will describe existing techniques attempting to meet the challenges of imaging the upper airway during speech and examine the remaining hurdles.

Static imaging of the vocal tract

The excellent soft tissue contrast and lack of ionising radiation have made MR imaging a popular research tool in analysing the