



Tooth visualization in vowel production MR images for three-dimensional vocal tract modeling

Ju Zhang^a, Kiyoshi Honda^a, Jianguo Wei^{b,*}

^a School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

^b School of Computer Software, Tianjin University, Tianjin 300350, China

ARTICLE INFO

Keywords:

Dental pulp
Volume-based landmark
Magnetic resonance imaging
Three-dimensional vocal tract modeling
Tooth visualization
Vowel production

ABSTRACT

Teeth are almost invisible in magnetic resonance imaging (MRI) because they lack free protons to magnetically react. In MRI-based studies on the vocal tract, the teeth must be visualized on the volume data obtained during articulation. To do so, varieties of techniques have been proposed, either covering the teeth by opaque materials or obtaining tooth images followed by their superimposition. In this article, a new method was proposed to visualize the teeth in vowel production MR images for the application of three-dimensional (3D) vocal tract modeling. 3D upper and lower jaw with the teeth was first extracted and reconstructed from static 3D-MRI data acquired during a simple ‘tooth imaging’ posture with minimal time and effort. The extracted 3D jaw with the teeth was superimposed onto the vowel production MRI volume three-dimensionally by using the dental pulps as volume-based landmarks to minimize fitting errors due to varied head positions across the scans. The effectiveness of the proposed method was demonstrated not only by the subjective opinions but also by the objective evaluation. The results show that the teeth are successfully and accurately superimposed onto the vowel production MR images. Also, the reconstructed 3D vocal tract models are observed with the bilateral interdental spaces after tooth superimposition. The proposed method solves the MRI-specific problem of the lack of tooth images and contributes to accurate 3D vocal tract measurement and reconstruction.

1. Introduction

Speech sounds with the phonetic quality result from the air column resonance in the vocal tract (Redford, 2015). Studies of vocal tract resonance patterns associated with vowels have long relied on 2D visualization of the vocal tract outline obtained by X-ray observation (Chiba and Kajiyama, 1942; Fant, 1960; Welch et al., 1989). Recent advances in digital imaging techniques have become available to acquire 3D shape information about the vocal tract volume by using volumetric imaging, such as magnetic resonance imaging (MRI) (Rokkaku et al., 1986; Baer et al., 1987) and computed tomography (CT) (Johansson et al., 1983; Perrier et al., 1992).

Among these techniques, MRI is a powerful imaging tool that has been extensively used in speech production studies (Baer et al., 1991; Story et al., 1996; Alwan et al., 1997; Shinagawa et al., 2005; Narayanan et al., 1995, 2014). The advantages of MRI include the ability to non-invasively obtain the entire vocal tract at a high spatial resolution with no hazardous exposure to radiation (Hardcastle et al., 2010). On the basis of MRI measurement, vocal tract acoustics associated with vowels have been widely investigated by many researchers

from the 3D vocal tract shape (Moore, 1992; Dang and Honda, 1997; Honda, 2004; Kitamura et al., 2005, 2006; Takemoto et al., 2006a; Delvaux and Howard, 2014).

Certainly, MRI also has a few drawbacks with respect to measuring the vocal tract. The major one is the invisibility of the teeth: they are as transparent as the air in MR images due to calcified structures having little mobile hydrogen to be detected by MRI. Accordingly, it is difficult to observe the relative positions of the tongue and teeth during speech production. Also, without tooth images, the oral cavity size could be easily overestimated in extracting vocal tract area functions (Story et al., 1996; Takemoto et al., 2004). Since the teeth form partial boundaries of the vocal tract, and the interdental spaces add unique characteristics to speech sounds, vocal tract visualization with the teeth in MR images is definitely necessary.

1.1. Related work

Various different approaches have been attempted to visualize the teeth in MR images, which are broadly classified into three representative categories.

* Corresponding author.

E-mail addresses: juzhang@tju.edu.cn (J. Zhang), khonda@sannet.ne.jp (K. Honda), jianguo@tju.edu.cn, jianguo_fr@163.com (J. Wei).

The first category is the dental cast scanning. Yang et al., (1995) crafted the dental impressions of the subject using the plaster. Coronal MR images of the dental impression placed in the water were taken at 1-mm intervals by T1 emphasis gradient echo high-speed scan using superconductive MRI system Sigma Advantage (GE Corp.). The 3D teeth were reconstructed from the boundaries extracted from the MR images. Hasegawa-Johnson et al., (2003) also made the dental cast of each participant and submerged them in water. Coronal and axial MR image stacks were generated by a Sigma 1.5 Tesla scanner (GE Corp.) with a 3-mm slice thickness for obtaining the 3D tooth shape. However, as reported in the previous studies (Kitamura et al., 2011; Ng et al., 2011; Ventura et al., 2014), foams produced on the dental cast may cause artifacts in MR images of the teeth.

The second category is using visible mouthpieces to cover the teeth. Wakumoto et al. (1997) developed the plates for upper and lower dental crowns from two-layered thermoforming materials. The contrast medium for MRI was enclosed and sealed in the plate, and the dental shape was extracted from the MR images collected by the methods of spin echo (SE) and turbo-FLASH using a SMT100GUX (Shimadzu Corp.). Since the dental plate is a few millimeters thick invisible layer, it is difficult to accurately define the air-tooth boundary. Kitamura et al. (2011) formed a dental mouthpiece made of the thermoplastic elastomer. The participant holding the dental mouthpiece in the mouth lay supine in a 3T MRI scanner (MAGNETOM Verio 3T, Siemens AG). The tooth shape in three dimensions was measured from the sagittal MR images acquired by the sequence of volume interpolated gradient breath-hold examination (VIBE). This method needs two MRI acquisitions (anterior-to-posterior and posterior-to-anterior) to compensate for the chemical shift artifact for volume data reconstruction. Ng et al. (2011, 2012) made a customized clear retainer filled with the jelly and performed a T1-weighted turbo-spin-echo (TSE-T1) sequence to visualize the teeth. Since only the incisor contours are defined, the 3D tooth shape cannot be obtained.

Use of liquid contrast medium in the oral cavity is the third category. Olt et al. (2004) instructed their volunteers to fill their mouth with the water. The dental structures were clearly distinguished through contrast with the surrounding fluid in combination with a 3D FLASH sequence. Takemoto et al. (2004) requested their subjects to hold the blueberry juice in the mouth as an oral contrast medium in a clinical 1.5T MRI scanner (Shimadzu-Marconi). The upper and lower teeth with the surrounding bony structures were visualized in MR images by fast spin echo (FSE) scan sequences. In these two methods, the subjects have to keep the prone position holding the contrast medium for a long data acquisition time. This is not comfortable for the subjects, and the artifacts may be caused by contrast fluid flow.

After the tooth visualization in MR images, the regions of teeth with the surrounding bony structures need to be segmented for the purpose of 3D shape measurement and reconstruction. MR image segmentation has long been a question to the researchers in the field of medical imaging. The gold-standard approach for accurate and robust segmentation is to trace the object boundary manually. However, such manual segmentation is weakened by shortcomings, including labor-intensive, reproducibility errors, operator fatigue and bias. A number of semi- or fully automatic algorithms have been developed (Li, 2009; Balafar et al., 2010; Heimann and Delingette, 2010; Guo et al., 2017). Several methods were proposed to segment the vocal tract and specific articulator (i.e. tongue) from 2D MR images or static 3D MR images (Bresch and Narayanan, 2009; Peng et al., 2010; Vasconcelos et al., 2011; Raeesy et al., 2013; Harandi et al., 2014; Javed et al., 2016). Despite the relative success of the above techniques, the segmentation of tooth MR images faces numerous challenges: The tooth boundaries have many cartographic changes with some tissues inside the tooth which leads to appearance of inner edges that make segmentation more troublesome. The tooth shapes are varied in different slice and across different speakers, which makes the statistical shape modeling methods not be accurate enough. The upper jaw with the teeth connects other orofacial

regions (e.g. varied nasal cavities). This results in the region-based methods could not suffice to delineate the object of interest. Therefore, the existing segmentation techniques may not be efficient for the challenges above, and so far the solutions presented in the previous literature for this task still largely rely on a manual delineation to guarantee the accuracy (Takemoto et al., 2004; Kitamura et al., 2011).

For the purpose of speech research, such as observation of articulatory movement and measurement of vocal tract shape, the obtained teeth are required to be superimposed onto the MR images during speech production. Ng et al. (2011, 2012) and Nunthayanon et al. (2015) superimposed the upper and lower incisor boundaries onto sequential 2D images of MRI movie using the landmarks along the cranium in the mid-sagittal plane. After the superimposition, the spatial-temporal relationship among articulators was evaluated during Japanese fricative and plosive articulation. Ventura et al. (2009) conducted the tooth superimposition using manual editing and pasting in the mid-sagittal plane for 3D vocal tract model reconstruction during the production of European Portuguese. Takemoto et al. (2004) superimposed the teeth via a 3D transformation by positional matching of point-based landmarks and measured vocal tract shapes during Japanese vowel production. Nevertheless, a notable problem is generally ignored among previous 2D and 3D tooth superimposition methods: the subject's head positions during tooth and articulation scanning can be varied in the 3D space. Consequently, tooth superimposition with rotation and translation only in the mid-sagittal plane is inadequate. For the 3D transformation, the use of the point-based landmarks may cause sampling errors within the slice thickness due to different slice locations of the head.

1.2. Overview

In the present study, we proposed a new method to visualize the teeth in MR images for the application of 3D vocal tract modeling. To achieve this goal, images of the teeth with the surrounding bony structures were first extracted from static 3D-MRI data acquired during a simple 'tooth imaging' posture with no need of subject's extra effort and discomfort. Subsequently, 3D upper and lower jaw with the teeth was reconstructed and then superimposed onto the target vowel production MRI volume data. Tooth superimposition was implemented in 3D space by using dental pulp volume-based landmarks to minimize the fitting errors caused by varied subject's head positions during tooth and speech production scanning. The effectiveness of the proposed method was examined not only by the subjective opinions but also by the objective evaluation. The goal of the study is set at the successful and accurate visualization of the teeth in vowel production MR images. To attain the goal, 3D vocal tract shapes during vowels were reconstructed after tooth superimposition, with the interdental spaces, i.e., the spaces between the upper and lower teeth.

The rest of the paper is structured as follows: Materials are introduced in Section 2. The methods and the details of each module are presented in Section 3. Section 4 shows the validations of the proposed method. Section 5 discusses and concludes the work.

2. Materials

2.1. Participants and speech tasks

Compared with the previous studies (Takemoto et al., 2004; Kitamura et al., 2011; Ng et al., 2011), more speakers, one 27-years-old male (WS) and two 28-years-old females (CR, LH), were requested as the subjects to validate the generality of the proposed method. They are all native speakers of Chinese with no history of speech disorders and previous and current jaw disease. All of them grew up in northern China and speak Mandarin without dialectal deviation.

The interdental spaces, enclosed by the upper and lower teeth and the cheek tissue, always move with the main vocal tract depending on

Download English Version:

<https://daneshyari.com/en/article/6960764>

Download Persian Version:

<https://daneshyari.com/article/6960764>

[Daneshyari.com](https://daneshyari.com)