Cine Magnetic Resonance Imaging With Simultaneous Audio to Evaluate Pediatric Velopharyngeal Insufficiency

Amanda L. Silver, MD; Katherine Nimkin, MD; Jean E. Ashland, PhD, SLP; Satrajit S. Ghosh, PhD; Andre J. W. van der Kouwe, PhD; Matthew T. Brigger, MD, MPH; Christopher J. Hartnick, MD, MSEpi

Objective: To develop a protocol linking cine magnetic resonance (MR) imaging to simultaneously acquired audio recordings of specific phonatory tasks to evaluate velopharyngeal insufficiency (VPI) in children.

Design: Institutional review board–approved development and application of a novel dynamic cine MR imaging protocol linked to simultaneously recorded audio.

Setting: A tertiary care multidisciplinary pediatric airway center.

Participants: Three healthy adult volunteers and 5 pediatric volunteers (age range, 9.3-18.9 years; mean age, 12.4 years) from the multidisciplinary pediatric airway center with VPI who previously had undergone nasopharyngoscopy, videofluoroscopy, or both.

Interventions: Cine MR imaging with simultaneously acquired audio files was performed in 3 adult volunteers to optimize the protocol and then in 5 pediatric volunteers meeting the inclusion criteria.

Main Outcome Measures: High-resolution cine MR images with clear intelligible audio recordings of specific phonatory tasks.

Results: Using 3 healthy adult volunteers, a cine MR imaging VPI protocol was developed that links simultaneously acquired cine MR images to audio recordings of specific validated phonatory tasks. Five school-aged children with VPI from our multidisciplinary pediatric airway center were then enrolled and underwent cine MR imaging using this protocol. The cine MR images and audio recordings acquired were of sufficient diagnostic quality to evaluate VPI closure patterns in school-aged children with VPI.

Conclusion: Cine MR imaging linked to audio is a quick, safe, and well-tolerated dynamic diagnostic imaging tool that may eventually have the potential to guide more precisely the selection and application of surgical techniques for VPI.


Velopharyngeal Insufficiency (VPI) refers to incomplete closure of the muscular valve separating the oral and nasal cavities and can lead to stigmatizing speech with compensatory grimacing; difficulty eating, drinking, or both; diminished performance in school; and decreased quality of life.1,2 Anatomically, VPI refers to failure in the posterosuperior movement of the soft palate (or velum), the medial movement of the lateropharyngeal walls, or a combination of both.1,3,4 As a result of this failure, air escapes through the nose during the production of oral pressure phonemes, causing a resonance disorder of speech.3

The etiology of VPI includes congenital conditions (eg, cleft palate) and acquired causes (eg, adenoidectomy). Considered by some to be a purely anatomical condition, research indicates that the true cause of VPI is likely multifactorial, with some combination of anatomical factors and neuromotor dysfunction. Cleft palate, which occurs in 1 of 625 to 762 live births, is the leading cause of VPI.5,6 Congenital conditions associated with VPI include velo-cardio-facial syndrome (22q11 syndrome), Down syndrome, muscular dystrophies, myasthenia gravis, and Möbius syndrome.7 Acquired causes include neuromotor (ie, tumors involving the vagus nerve, brainstem strokes, or traumatic brain injury) or those resulting from surgery, with an estimated 1 in 1300 to 10 000 patients developing VPI following adenoidectomy.3

The accepted diagnostic standard includes a combination of perceptual speech analysis by trained speech-language pathologists in conjunction with nasopharyngoscopy (NP), videofluoroscopy (VF), and...
Typically, 1 or 2 (but not all) of these diagnostic modalities are used. Speech and resonance evaluations involve assessment of articulation, nasal emissions, laryngeal and nasal resonance, and oral motor function. Both NP and VF provide useful clinical information; however, each assessment modality has specific drawbacks that render clinical judgment imperative when choosing between 1 or both of these modalities.

Nasopharyngoscopy offers a direct top-down or bird’s-eye view of velopharyngeal closure (or gap) when combined with specific phonatory tasks and may be performed in the office using topical anesthesia. However, this procedure is invasive and requires patient cooperation to tolerate the endoscope and to simultaneously perform the necessary speech tasks with the endoscope in the proper position. Therefore, age may be a factor in the success of this procedure, as young children may tolerate the examination but may not speak volitionally when asked, thereby limiting the usefulness of the examination. In addition, patients may not tolerate a successive examination. There may be wide-angle distortion or glare, and bulky adenoid tissue may obscure the extent of closure. Even in the best of circumstances, evaluation and quantification of the velopharyngeal gap are limited to 2 dimensions (anteroposterior and lateral) at the most superior level of closure.

Videofluoroscopy is a radiographic modality that enables the clinician to identify the height of velopharyngeal closure with reference to vertebrae and may determine involvement, if any, of Passavant ridge. Three views typically are obtained, namely, lateral, frontal, and base (or Towne projection). Unfortunately, VF subjects the patient to ionizing radiation, which in part limits the time to capture images. This method also requires patient cooperation to perform necessary speech tasks in a timely manner after barium has been instilled into the patient’s nose, a procedure that is often irritating and uncomfortable. It can be challenging to obtain a good base view, and there may be multiple shadows that complicate the interpretation of VF imaging in patients who have undergone pharyngeal flaps. Because of the discomfort and risks (notably ionizing radiation) to the patients, VF use has declined in favor of NP.

Therefore, there is no true criterion standard diagnostic modality that not only is well tolerated and safe but also provides evaluation of velopharyngeal closure in all planes and that has the potential to guide the choice of surgical technique by linking images of muscle movement during phonation of well-established phonatory tasks. In addition, there is no accepted or proven best practice algorithm that guides diagnostic modality selection. Dynamic visualization and measurements of the anatomical defect leading to VPI in a patient should enable individualized surgical planning by allowing the surgeon to precisely target the specific underlying muscular defect that is leading to VPI in each individual patient.

First described in 1999, cine magnetic resonance (MR) imaging is a dynamic modality that offers superior visualization of soft tissues without exposure to ionizing radiation. Whereas image quality in standard MR imaging is dependent on a motionless individual, the cine MR images are captured during movement (eg, heartbeats or speech) and can then be replayed in real time like a movie. Cine MR imaging has been used to evaluate speech in healthy individuals and in patients with VPI and (separately) to evaluate articulatory movement in patients with cleft lip and palate. Although several authors have reported using dynamic MR imaging to obtain quantitative measurements of the levator veli palatini muscle at rest and during speech in healthy individuals or in patients after cleft palate repair, we are unaware of any reports linking audio recordings to cine MR imaging in patients having VPI with a goal of defining the specific anatomical defects. Without linking simultaneously acquired audio recordings to the cine MR images, it may be difficult to distinguish speech from swallowing or to evaluate the pattern of VPI based on the specific phonatory task. A movie that combines simultaneously acquired audio with cine MR images can be replayed as often as necessary by any member of the multidisciplinary team caring for patients with VPI.

In this pilot study, we sought to demonstrate that cine MR imaging linked to audio recordings is a safe and valid technique that is well tolerated by school-aged children for the evaluation of VPI. We developed a cine MR imaging protocol that links cine MR images to audio recordings of specific phonatory tasks captured simultaneously in a movie format. Our goal was to develop a quick, safe, and well-tolerated dynamic diagnostic imaging tool to potentially better guide the selection of surgical techniques.

METHODS

After obtaining institutional review board approval, 3 healthy adult volunteers underwent imaging during phonation to develop our cine MR imaging VPI protocol. Images were obtained on a 3.0-T MR imaging system (Siemens Healthcare Diagnostics, Deerfield, Illinois), the same machine used for pediatric MR imaging at our institution. In the development of our MR image acquisition protocol, we sought to maximize image quality while decreasing image acquisition time to increase pediatric patient tolerance and compliance. Using integrated parallel acquisition techniques technology, we were able to obtain 50 diagnostic-quality cine images in sequences lasting 26 to 27 seconds. Study participants were placed in the supine position in a head-neck coil and were awake during the entire examination. A parent was allowed to sit with the patient in the imaging room, provided that the parent was able to undergo MR imaging. After 3-dimensional localizers, sagittal T1-weighted and short tau inversion recovery (STIR) sequences were obtained during quiet breathing for anatomical evaluation. Cine images then were obtained during phonation using 2-dimensional fast low-angle shot (FLASH) (gradient echo) sequences in 3 planes (6-mm section thickness, 5.0-millisecond repetition time, 1.94-millisecond echo time, and 8° flip angle). Sagittal images were centered over the area of the posterior oropharynx. Axial images were obtained at the level of VP closure, and coronal images were obtained at the center of the oropharyngeal lumen.

Because current MR imaging systems lack integrated audio hardware and software, we developed an audio circuit. The hardware included a combination of inexpensive audio equipment and items readily available in the hospital, such as oxygen tubing, tape, and plastic foam cups. A plastic foam cup, held by the participant at his or her mouth during the cine MR imaging,
was connected to oxygen tubing using hospital tape. The other end of the oxygen tubing was taped to a microphone. The microphone wiring then was fed through the wave guide, an existing hole in the wall between the room with the MR imaging system and the MR imaging technician’s room. This ensured that the microphone, which contained a trace amount of ferrous material, was kept at a sufficient distance from the MR magnet. In the technician’s room, the microphone wiring was connected to a USB preamplifier (http://www.m-audio.com/products/en_us/MobilePreUSB.html) that in turn was connected via a standard USB cable to a laptop PC.

We downloaded free open-source software for recording and editing audio files (http://audacity.sourceforge.net) to capture and save the audio recordings at the time of each cine MR imaging. During image acquisition, audio was simultaneously recorded of each participant repeating 1 of the following 2 target phrases: “pick up the puppy” or “Suzy has shoes.” A speech pathologist (J.E.A.) chose the target phrases because they contain both high-pressure and low-pressure nonnasal consonants. The selected phrase was repeated during the entire duration of image capture to ensure consistent and reproducible results. To ensure that the entire audio file for each imaging sequence was recorded, the study participant was asked to begin speaking before the start of image or audio acquisition. We began recording the audio files just before the start of each MR imaging sequence. Immediately after each sequence was completed, the participant was instructed to stop speaking, and both the audio and imaging files were saved and clearly labeled, with the duration of the sequence clearly noted to later be able to combine the correct audio file with the correct MR imaging sequence.

Once the cine MR imaging VPI protocol was optimized, we enrolled 5 school-aged volunteers with VPI. To be eligible, volunteers had to be patients in the multidisciplinary pediatric airway center at our institution and had to have already undergone NP, VF, or both. The study participants had to be able to undergo MR imaging (ie, they could not have any implanted ferrous devices). The participants and their parents signed study assents and consents, respectively. Each pediatric volunteer spent approximately 30 minutes awake and in the supine position in the imager. After 3-dimensional localizers, sagittal T1-weighted and STIR sequences were obtained; cine MR imaging sequences in the sagittal, coronal, and axial planes (each lasting <30 seconds) were procured. Six sequences were recorded so that sagittal, coronal, and axial images were obtained as participants simultaneously repeated 1 of 2 target phrases. The study participants were then removed from the MR imager, and the cine MR images were saved and transferred to a CD.

After each audio file was saved as a separate. WAV file, it was cropped to contain only the portion corresponding to the time of image acquisition. Each audio file (.WAV file) then was combined with the corresponding MR image file (AVI file) using commercially available software (Windows Movie Maker; Microsoft, Redmond, Washington), resulting in 6 movies combining audio and video for each participant (ie, sagittal, coronal, and axial movies of the participant repeating “pick up the puppy” and sagittal, coronal, and axial movies of the participant repeating “Suzy has shoes” ) (video; http://www.archoto.com).

RESULTS

Three healthy adult volunteers underwent cine MR imaging as they were simultaneously being recorded speaking the phrases “pick up the puppy” and “Suzy has shoes.” The protocol was established and optimized over a total of 8 sessions. Five school-aged children with VPI then underwent cine MR imaging using the same protocol. The study participants ranged in age from 9.3 to 18.9 years (mean age, 12.4 years). Demographics and pertinent medical and surgical histories are listed in the Table. We were able to specifically and directly identify the velopharyngeal closure pattern and areas of insufficiency. Multiplanar imaging allowed assessment of lateral wall motion as well as palatal movement in axial, coronal, and sagittal planes. This is critical because, as suggested by our preliminary data, “normal” patients (ie, patients without clinical evidence of VPI during speech) at times may have incomplete VP closure; similarly, patients with clinical evidence of VPI at times achieve complete VP closure (eg, when swallowing). These findings speak to the anatomical and physiological complexities of speech and swallowing, and a multiplanar diagnostic imaging modality may help to better understand and treat disorders of speech and swallowing, such as VPI.

Because of the invasive nature, exposure to radiation (in VF), and limitations of the current diagnostic modalities for VPI (NP and VF), we sought to develop a safe, noninvasive, and well-tolerated diagnostic imaging tool for the evaluation of VPI in school-aged children. Our cine MR imaging VPI protocol provides high-resolution diagnostic-quality cine MR images linking relevant upper airway anatomy to simultaneously acquired validated phonatory tasks without exposing the individual to ionizing radiation. Our protocol was established and optimized using healthy adult volunteers and was then applied to evaluate 5 school-aged volunteers with VPI. Multiplanar imaging allowed assessment of lateral wall motion, as well as palatal movement in axial, coronal, and sagittal planes. We were able to specifically and directly identify the velopharyngeal closure pattern and area(s) of insufficiency.

The initial challenge of this project was to develop a method to capture audio recordings simultaneously with MR imaging. Although current MR imaging systems are capable of capturing cine MR imaging, they lack the integrated hardware and software necessary to capture simultaneously acquired audio recordings. Therefore, we needed a freestanding laptop PC, as well as audio hardware and software, to accomplish our goal. We combined readily available recording equipment (MobilePre USB preamplifier and a microphone) with free open-source software (Audacity), generic hospital equipment (oxygen tubing, tape, and plastic foam cups), and a standard laptop PC.

Although we are able to capture phonatory tasks simultaneously in recordings of adequate quality, there are limitations to our technique. The amount of ambient or background imaging system noise recorded during each sequence depended on the placement of the plastic foam cup with respect to the participant’s mouth, as well as movement of the cup during the sequence. Because this movement is irregular and unpredictable, it cannot subsequently be subtracted easily from the audio tract. Our
Table. Demographics and Medical and Surgical Histories of Pediatric Patients

<table>
<thead>
<tr>
<th>Patient No./Sex/ Age at Cine MR Imaging, y</th>
<th>Relevant Examination Findings</th>
<th>Medical History</th>
<th>Surgical History</th>
<th>Cine MR Imaging Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/F/9.7</td>
<td>NP showed central gap between soft palate and pharynx due to a short velum</td>
<td>Multiple ear infections, eczema</td>
<td>. . .</td>
<td>Marked enlargement of palatine and lingual tonsils and moderate adenoid prominence; intermittent touch contact of soft palate with posterior pharynx at the level of the adenoids during phonation, persistent small gap in all 3 planes with occasional touch closure; enlarged palatine tonsils are positioned below the soft palate and do not appear to interfere with velopharyngeal closure</td>
</tr>
<tr>
<td>2/F/11.8</td>
<td>NP showed small central gap, VF showed mild soft palate hypotonia</td>
<td>Turner syndrome, thoracic scoliosis, conductive hearing loss in both ears</td>
<td>Posteropharyngeal wall rolled flap pharyngoplasty (13 mo before cine MR imaging)</td>
<td>T2-weighted bright signal in the posteropharyngeal wall at the level of the soft palate likely reflects residual adenoidal tissue and postoperative changes; prominent lingual tonsils; with Suzy phrasing, there is no touch contact noted between the palate and posterior pharyngeal wall with a small central gap noted on axial sequences; with puppy phrasing, there is improved contact between the soft palate and posterior pharyngeal wall without significant gap noted on axial sequences; normal lateral wall motion is demonstrated during phonation</td>
</tr>
<tr>
<td>3/M/9.3</td>
<td>NP showed small central gap, VF showed good movement of soft palate and lateropharyngeal walls</td>
<td>Submucous cleft palate, attention-deficit/hyperactivity disorder, chronic ear infections, speech services since age 2 y</td>
<td>Myringotomy tube placement 3 times</td>
<td>Moderate tonsillar and adenoidal tissue; no significant velopharyngeal insufficiency demonstrated on sagittal images with puppy phrasing; decreased contact of the soft palate with the posteropharyngeal wall with Suzy phrasing; small central gap seen on axial images, especially with Suzy phrasing; normal lateropharyngeal wall motion</td>
</tr>
<tr>
<td>4/M/18.9</td>
<td>NP showed small central gap, VF showed short soft palate, good movement of lateral walls, mild bulge of Passavant bar</td>
<td>Mild bifid uvula, submucosal cleft, speech therapy from ages 3-15 y</td>
<td>Tonsillectomy and adenoidectomy at age 6 y, myringotomy tube placement, sphincter pharyngoplasty (4 mo before cine MR imaging)</td>
<td>No residual adenoidal or tonsillar tissue; persistent small gap between the soft palate and posterior pharyngeal wall with Suzy and puppy phrasing; on axial images, the gap appears larger on the left and right; normal symmetric lateral wall motion during phonation</td>
</tr>
<tr>
<td>5/F/9.7</td>
<td>NP showed marked foreshortening of palate with broad velopharyngeal insufficiency</td>
<td>Bilfided uvula, submucosal cleft, short palate, prior genetic evaluation negative for velo-cardio-facial syndrome</td>
<td>Palatoplasty at age 5 y at outside hospital</td>
<td>Soft palate appears foreshortened; minimal residual adenoidal tissue, cephalad to the area of the persistent gap; there is a persistent gap during phonation between the posterior soft palate and the posteropharyngeal wall in all 3 planes, the gap is more prominent during puppy phrasing compared with Suzy phrasing, and the gap is noted at the level of the anterior arch of C1 and is slightly larger on the left than the right on axial images; the gap measures approximately 4 mm in anteroposterior dimension</td>
</tr>
</tbody>
</table>

Abbreviations: Ellipses, not applicable; MR, magnetic resonance; NP, nasopharyngoscopy; VF, videofluoroscopy.

Another challenge involved postacquisition processing of the audio and imaging files. The manually recorded audio tracks of phonatory tasks during image acquisition needed to be combined with the cine MR images. Although our current image acquisition protocol yields high-resolution diagnostic-quality images captured in less than 30 seconds, the protocol does not capture an integer number of frames per second (but rather 1.85 or 1.92 frames per second). In combination with the inherent imprecision of manually cropping the audio files, this leads to a slightly imperfect linking of the audio and imaging files that seems to become more apparent over the course of each movie or sequence. The inherent imprecision of linking the image and audio files remains a limitation of this study. Ongoing efforts are being made to develop an image acquisition protocol that captures an integer num-

©2011 American Medical Association. All rights reserved.
number of images over a short duration (<30 seconds) while maintaining or even improving image quality. Ultimately and ideally, audio hardware and software would be integrated in the MR imaging system to facilitate audio capture and to streamline the linking of audio and image files. Such integrated equipment may obviate the need to manually link the audio and image files and eliminate this current source of imprecision.

Another potential source of error stems from the method in which cine MR images are acquired. In our current protocol, there is a degree of interpolation between the cine MR images acquired that is necessary to generate the movie format. Although the duration of this interpolation is short, it is theoretically possible that instantaneous moments of touch closure between the velum and posterior wall, for example, are missed. Therefore, our technique may yield false-positive results by diagnosing VPI when, in fact, full closure has occurred almost immediately. However, our simultaneously recorded audio tracts should help prevent this error because we are able to assess each participant’s speech clinically for evidence of VPI. Nevertheless, neither the exact duration of VP closure or the optimal closure pattern (ie, a complete seal or touch closure) is clearly known nor whether closure over a small segment or along a broader zone is necessary. These factors should be clarified on an individualized and normative basis.

There are several other potential limitations of our protocol. The availability and cost of cine MR imaging, particularly as it compares with NP or VF, are to be considered. In our institution, the approximate total cost of NP is $400 (physician charge) and of VF is $1800 (radiologic professional charge, radiologic technician charge, and speech-language pathologist charge) compared with $5000 for cine MR imaging (radiologic professional charge and radiologic technician charge). Study participants must be able to undergo MR imaging (ie, have no implanted ferrous objects) and must be able to tolerate lying in the supine position in the MR imager for approximately 30 minutes and to simultaneously comply with directions during image acquisition. Children who are claustrophobic or frightened by the noise of the MR imaging system may not tolerate this protocol. In our study, we sought to maximize patient comfort and compliance by allowing a parent to sit with the patient in the MR imaging room while the child underwent image acquisition, provided that parent could undergo MR imaging as well. Furthermore, although the effect of body position (ie, upright or supine) on speech has been studied, this variable requires further investigation.

This was a pilot study designed to establish a novel protocol for cine MR images linked to simultaneously recorded phonatory tasks in a safe, well-tolerated, and feasible manner. Future work is necessary, not only to refine the protocol as already discussed but also to further the understanding of VPI. Most important, normative data are needed. A larger series of healthy patients undergoing cine MR imaging is necessary to better define normal velopharyngeal closure patterns using MR imaging. A larger series of patients with VPI then needs to be studied to better understand the MR imaging appearance of common anatomical abnormalities that lead to VPI. It is likely that a longitudinal study of children with VPI who undergo serial NP, VF, and cine MR imaging linked to simultaneously acquired audio before and after intervention (ie, speech therapy, surgery, or both) could contribute to normative data. Furthermore, a large-scale formal quantitative comparison of NP, VF, and cine MR imaging needs to be performed.

Last, because of the neuromuscular complexity of speech, functional MR imaging combined with cine MR imaging linked to audio eventually may provide an even more targeted understanding of normal and abnormal speech in patients with and without anatomical evidence of VPI. Such an imaging modality could help to define the extent to which VPI is a neuromuscular or anatonic process. By comparing such imaging of speech with imaging of swallowing, we may be able to uncover different signaling pathways that lead to velopharyngeal function, which may in turn lead to advances in speech therapy and in surgical techniques.

In summary, our protocol offers a quick, safe, and well-tolerated dynamic diagnostic imaging tool that links images to simultaneously acquired audio in school-aged patients with VPI. It remains to be determined whether cine MR imaging linked to simultaneously acquired audio can be used not only to increase the number of patients with VPI who may be evaluated but also to develop best practice clinical algorithms for the evaluation, diagnosis, treatment, and outcomes of patients with VPI. Future work is necessary to compare cine MR imaging with NP and VF, as well as to determine if this technique might facilitate the evaluation of pediatric patients with VPI and may more specifically and directly guide the choice and application of surgical techniques compared with the current diagnostic modalities for VPI.

Submitted for Publication: May 17, 2010; final revision received September 16, 2010; accepted November 9, 2010.

Author Affiliations: Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, and Department of Otolaryngology and Laryngology, Harvard Medical School (Dr Silver and Hartnick); Department of Radiology (Dr Nimkin) and Department of Speech and Language Pathology, Speech Communication Group (Dr Ashland), Massachusetts General Hospital, Harvard Medical School, Boston, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge (Dr Ghosh), and Athinoula A. Martinos Center for Biomedical Imaging and Department of Radiology, Massachusetts General Hospital, Charlestown (Dr van der Kouwe); and Department of Otolaryngology—Head and Neck Surgery, Naval Medical Center San Diego, San Diego, California (Dr Brigger).

Correspondence: Christopher J. Hartnick, MD, MS Epi, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, 243 Charles St, Boston, MA 02114 (Christopher_Hartnick@meei.harvard.edu).

Author Contributions: Dr Silver had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Silver, Nimkin, Ashland, and Hartnick. Acquisition of data: Silver, Nimkin, Ghosh, van der Kouwe, Brigger, and Hartnick. Analysis and interpretation of data: Silver and Hartnick. Drafting of the manu-
REFERENCES