Erik Bresch, Yoon-Chul Kim,
Krishna Nayak, Dani Byrd,
and Shrikanth Narayanan

Seeing Speech: Capturing Vocal Tract Shaping Using Real-Time Magnetic Resonance Imaging

Understanding human speech production is of great interest from engineering, linguistic, and several other research points of view. While several types of data available to speech understanding studies lead to different avenues for research, in this article we focus on real-time (RT) magnetic resonance imaging (MRI) as an emerging technique for studying speech production. We discuss the details and challenges of RT magnetic resonance (MR) acquisition and analysis, and modeling approaches that make use of MRI data for studying speech production.

MOTIVATION

From an engineer's point of view, detailed knowledge about speech production gives rise to refined models for the speech signal that can be exploited for the design of powerful speech recognition, coding, and synthesis systems. From a linguist's point of view, speech research may be conducted to address open questions in the areas of phonetics and phonology. These include: 1) what articulatory mechanisms explain the inter- and intrasubject variability of speech, 2) what aspects of the vocal tract shaping are critically controlled by the brain for conveying meaning and emotions, and 3) how does prosody affect the articulatory timing. From other research points of view, speech production is important to understand language acquisition and language disorders. All of these efforts require intimate knowledge of the speech generation mechanisms.

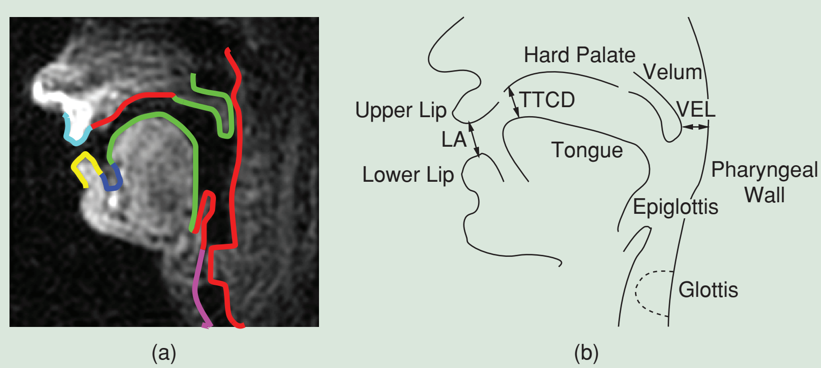
Different types of data are available to the speech researcher—from audio and video recordings of speech production to muscle activity data produced by electromyography, respiratory data from subglottal or interoral pressure transduction, and images of the larynx obtained through video laryngoscopy. While the vocal tract posture and movement can be investigated using a host of techniques summarized in Table 1 including X ray (microbeam), cinefluography, ultrasound, palatography, electromagnetometry (EMA), RT-MRI has a particular advantage in that it produces complete views of the entire vocal tract including the pharyngeal structures in a safe and noninvasive manner.

With RT-MRI, a midsaggital image of the vocal tract from the glottis (bottom) to the lips (left) can be acquired as illustrated in Figure 1(a). In this image, we can trace the air-tissue boundaries of the anatomical components that are of interest to the speech researcher and obtain a representation similar to Figure 1(b). These components, also known as articulators, are controlled by the brain during speech production and are used to change the shape of the vocal tract tube. With it, they also change the filter function for the excitation signal generated at the glottis and elsewhere along the airway. Hence the motion of the articulators shapes the sounds of speech and other human vocalizations.

The signal processing challenges when studying these using RT-MRI lie in the fast acquisition of high-quality RT MRI images including simultaneous noise-robust audio recording [1], the subsequent detection of the relevant features from each image, and the analysis and modeling of the time-varying vocal tract shape for the purpose of gaining deeper understanding of the underlying principles that govern the speech production process.

RT-MRI DATA ACQUISITION AND ANALYSIS

MR images are formed through the interaction between externally applied magnetic fields and nuclear magnetic spins (primarily ^{1}H). Hydrogen content within a region or slice of interest is excited using a radiofrequency magnetic



[FIG1] Example vocal tract MR image and tract variables. (a) Typical midsagittal real-time MR image with contours of interest. (b) Articulators and some sample vocal tract variables: lip aperture (LA), tongue tip constriction degree (TTCD), and velum aperture (VEL).

Digital Object Identifier 10.1109/MSP.2008.918034

field, and data is recorded while linear magnetic field gradients are oscillated. The recorded signals represent samples in the spatial Fourier transform domain, which is referred to as k-space, of the excited object. For RT speech applications, gradient echo imaging with short echo times provides the necessary speed while minimizing signal loss.

RT-MRI refers to the continuous acquisition of MR images with frame rates sufficient to capture the underlying motion or physiology of interest, typically 5–50 frames per second. Such systems are often interactive, requiring reconstruction and display of images with minimal system latency, typically 100–500 ms. This has become possible due to advances in imaging technology including rapidly fluctuating gradients, novel k-space trajectories, parallel imaging, as well as improvements in computer speed.

RT-MRI of human speech production [2] faces several important constraints. The magnetic susceptibility difference between air and human tissue creates local variations in resonance frequency, which affects recorded MRI signals. In addition, the required spatial and temporal resolution requirements are not known in advance and are expected to vary for different speech tasks. Finally, the study of the upper airway has not been the focus of much previous MRI development, and most equipment and imaging methods are not optimized for this region of interest.

SEQUENTIAL RT-MRI ACQUISITION, RECONSTRUCTION, AND EDGE DETECTION

Various MRI scan schemes are possible to collect the k-space data, such as two-dimensional (2-D) discrete Fourier transform (DFT), echo planar imaging (EPI), radial, or spiral readout trajectories, all of which have different advantages and disadvantages in terms of scan speed, off-resonance and motion artifact suppression, and signal-to-noise ratio. Studies have shown that sampling the k-space data along a spiral trajectory provides a good compromise and delivers reasonable results. However, the fact that the k-space samples do not fall on a Cartesian grid requires a modified reconstruction.

Assume the system consists of the blocks illustrated in Figure 2(a). Full sampling of the k-space data along a spiral trajectory is followed by gridding reconstruction, which yields an image on which we can perform edge detection to identify the contours of interest.

Gridding reconstruction is the most popular method for reconstructing spiral data [3]. It is fast in terms of the reconstruction time, and it produces images with reasonable reconstruction accuracy. In gridding reconstruction,

$$\hat{m}(x, y) = \frac{\mathrm{iFFT}_{2D}((M \cdot W) * R \cdot \mathrm{III})}{r(x, y)}, \tag{1}$$

where $\hat{m}(x, y)$ denotes a pixel value at (x, y) of a 2-D image after gridding reconstruction, M is the k-space data matrix, W denotes density correction factors, R is a convolution kernel, and III is a 2-D uniformly spaced Cartesian resampling function. To obtain $\hat{m}(x, y)$ the following steps are required:

1) Convolve the k-space data M, multiplied by the density correction

[TABLE 1] METHODS FOR ACQUIRING SPEECH PRODUCTION DATA OF THE VOCAL TRACT.

METHOD	PROS	CONS	COMMENTS
CT	• HIGH TEMPORAL AND SPATIAL RESOLUTION • CAPTURES PHARYNGEAL STRUCTURES • 3-D POSSIBLE	• EXPOSURE TO RADIATION	• RARELY USED IN SPEECH RESEARCH
EMA	• HIGH SPATIAL AND TEMPORAL RESOLUTION • 3-D	• PROVIDES SPATIALLY SPARSE POINT TRACKING DATA • CANNOT CAPTURE PHARYNGEAL STRUCTURES	• OFTEN USED IN SPEECH PRODUCTION STUDIES.
X RAY (MICROBEAM)	• HIGH SPATIAL AND TEMPORAL RESOLUTION • FLESH POINT TRACKING NOT POSSIBLE FOR PHARYNGEAL STRUCTURES	• EXPOSURE TO RADIATION • IMAGES SHOW ONLY A PROJECTION THROUGH VOLUME WHICH MAKES CONTOUR EXTRACTION DIFFICULT • X-RAY MICROBEAM EQUIPMENT NOT WIDELY AVAILABLE • SPATIALLY SPARSE DATA	• RARELY USED NOW IN SPEECH RESEARCH • EXISTING DATABASES ARE STILL BEING USED
ULTRASOUND	• HIGH TEMPORAL RESOLUTION • NON-INVASIVE, SAFE • GOOD AUDIO CAN BE OBTAINED SIMULTANEOUSLY	• NOISY IMAGES • DETECTS ONLY FIRST TISSUE-AIR BOUNDARY • NOT SUITABLE FOR ANTERIOR TONGUE TIP AND LIP IMAGING • DETECTOR IS IN CONTACT WITH JAW AND MAY AFFECT SPEECH PRODUCTION PROCESS	• USED PRIMARILY FOR TONGUE BODY IMAGING
MRI	• NON-INVASIVE, SAFE • CAPTURES PHARYNGEAL STRUCTURES • 3-D POSSIBLE • TAGGED MRI ALLOWS FLESH-POINT TRACKING	• RELATIVELY POOR SPATIAL AND TEMPORAL RESOLUTION • EXPENSIVE • LIMITED TO SUBJECTS WITHOUT MAJOR DENTAL WORK/IMPLANTS • SUPINE POSITION GENERALLY REQUIRED • SIMULTANEOUS AUDIO RECORDING DIFFICULT DUE TO SCANNER NOISE	• AN EMERGING TECHNIQUE FOR SPEECH RESEARCH

factors W, with a convolution kernel R.

2) Resample the result into a uniformly spaced grid III.

3) Compute a fast inverse Fourier transform of the data on the uniformly spaced grid.

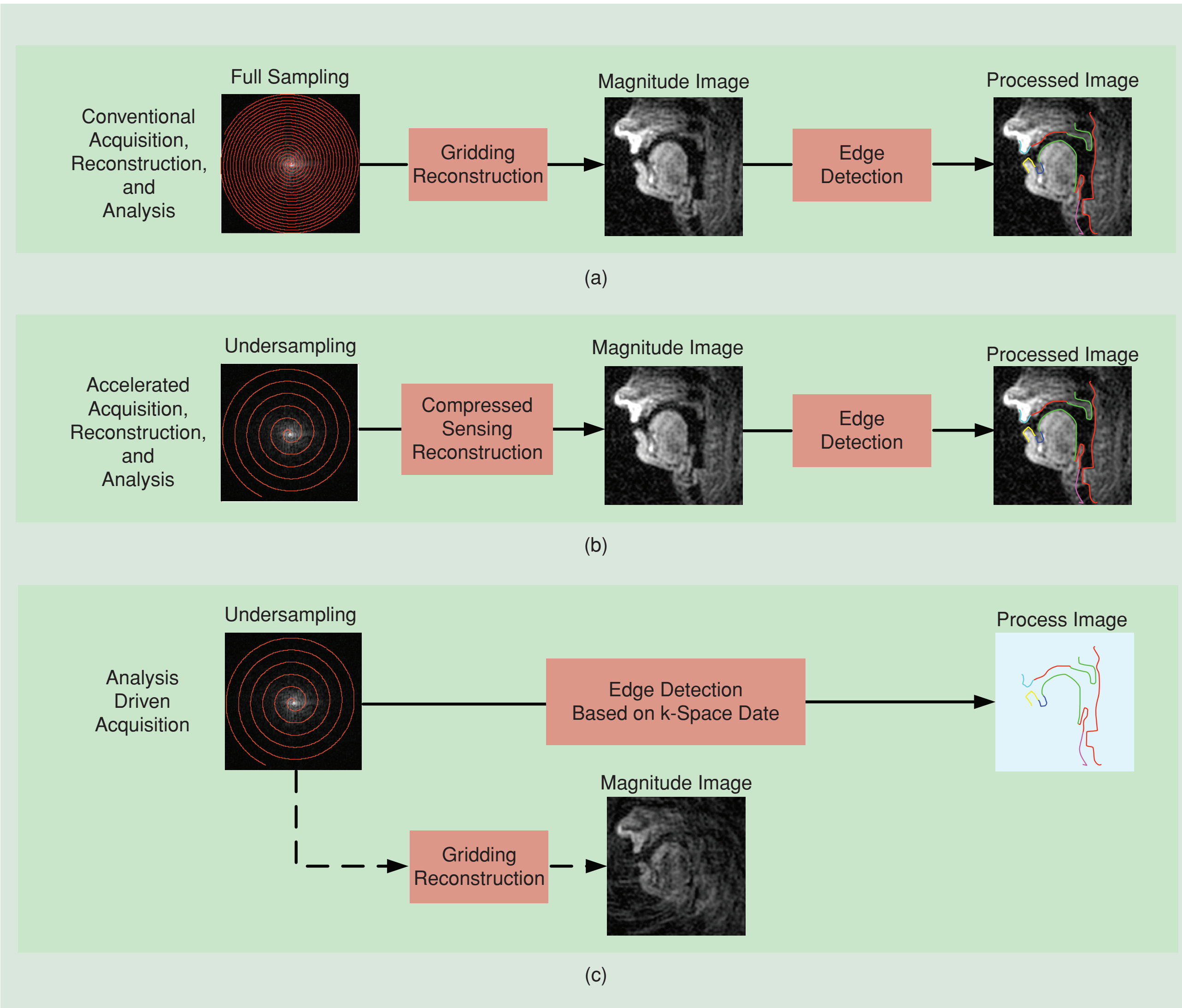
4) Divide the resulting image by a 2-D Fourier transform $r(x, y)$ of R to correct for attenuation effects due to the finite extent of R.

The outcome of the gridding reconstruction consists of a complex valued 2-D intensity image, whose absolute value $m(x, y) = |\hat{m}(x, y)|$ is commonly displayed and used for the subsequent edge detection process.

The edge detection in a given image can be formulated as an optimization process in which we search for the best set of parameters P of a contour $C(P)$ to align it with the intensity gradients, i.e., with the edges in the image [4]. To quantify the goodness of the fit of a candidate contour to the underlying image edges, an external energy measure E_{ext} is adopted, which is chosen here to be the line integral along the contour C over the negative image intensity gradient magnitude

$$E_{\text{ext}}(P) = -\int_{C(P)} |\nabla(h(x, y) * m(x, y))|\, ds \qquad (2)$$

An optional smoothing filter kernel $h(x, y)$ may be applied to mitigate the effects of noise in the image. To produce smooth contours an internal energy measure $E_{\text{int}}(C(P))$ can be used to measure and penalize the curvature, or roughness, of the candidate contour.

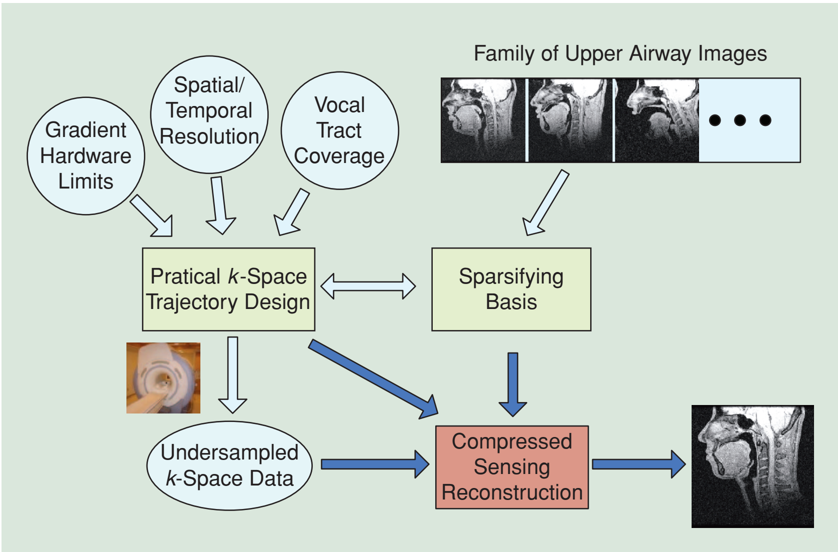


[FIG2] Data acquisition and analysis methodologies. (a) Conventional approaches acquire fully sampled spiral data, produce images using gridding reconstruction, and perform edge detection to extract vocal tract contours. (b) Accelerated spiral acquisition can be achieved by undersampling spiral trajectories but will produce aliasing artifacts in images when using gridding reconstruction. Advanced reconstruction techniques such as compressed sensing (shown), non-Cartesian paralell imaging, or temporal filtering, can in some cases produce alias-free images. Edge detection is performed to produce final vocal tract contours. (c) The analysis-driven approach considers only the final output (e.g., vocal tract contours) without requiring high quality intermediate data (e.g., magnitude image). This relaxes the imaging constraints and can allow even greater acceleration.

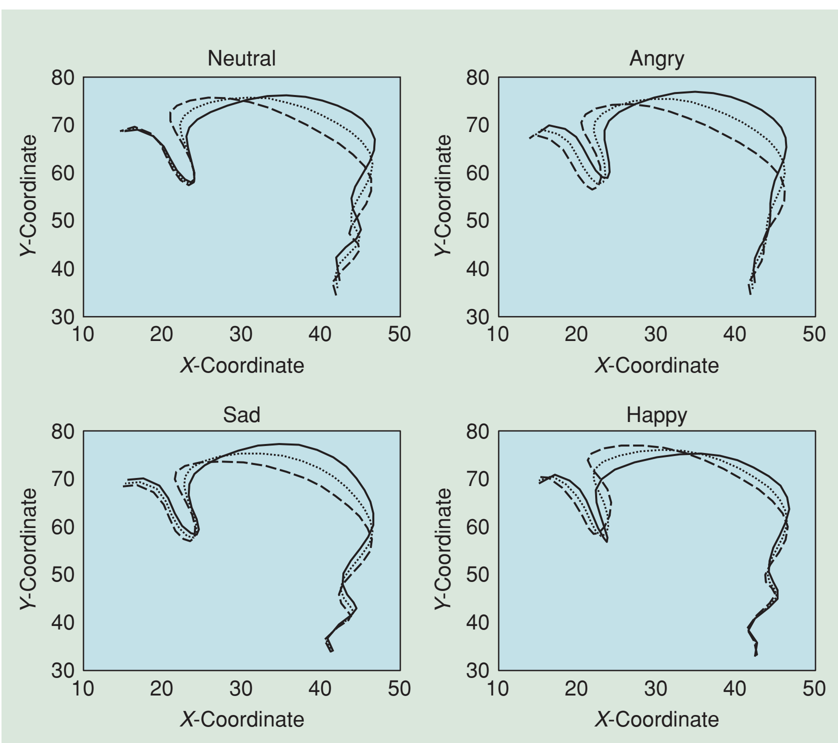
The contour detection process now consists of minimizing the sum of external and internal energies associated with the contour by adjusting the parameters that control the contour shape

$$\hat{P} = \underset{P}{\operatorname{argmin}} \left(E_{\text{ext}}(P) + E_{\text{int}}(P)\right). \quad (3)$$

The challenges here consist of how to a) select the appropriate contour descriptors that allow a straightforward computation of (2) and allow the capturing of the specifics of the upper airway contours with few parameters, b) design an algorithm to solve the nonconvex optimization problem in (3), and c) detect robustly the endpoint of the contour segments that correspond to the individual articulators. A solution to b) may rely on exploiting knowledge of the vocal tract anatomy and the deformation space of the individual anatomical components, and a solution to c) may consist of employing model-guided optimization. However, these challenges remain open for further study.



[FIG3] Compressed sensing reconstruction applied to real-time speech MRI. Two important considerations are time-efficient gradient waveform design (and its associated k-space sampling pattern) and the choice of the sparsifying basis.



[FIG4] Modulation of the tongue shape by the first PCA component for four different emotions (mean shape with dotted line).

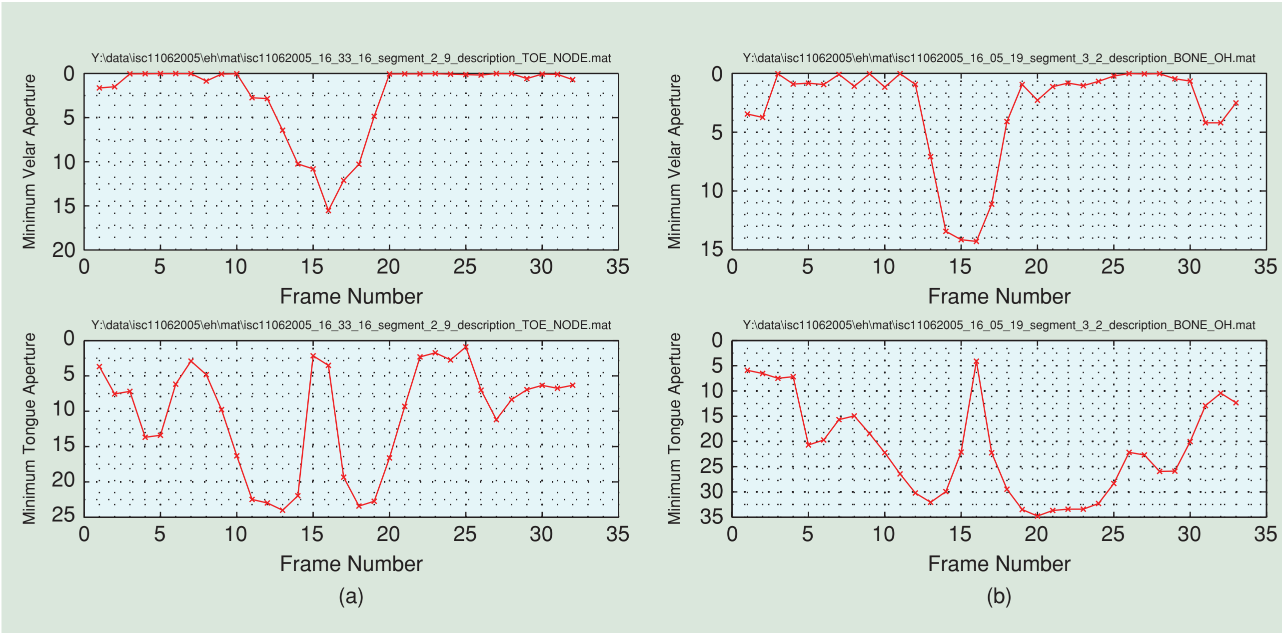
ACCELERATED ACQUISITION, COMPRESSED SENSING RECONSTRUCTION AND EDGE DETECTION

Currently achievable frame rates with the conventional RT-MRI techniques are sufficient for some speech research studies. However, for safety reasons in practice the frame rates cannot be further increased by accelerating the readout scans and hence the delivered MRI data rate.

Instead, to capture the fastest vocal tract shaping phenomena a viable solution is to employ accelerated methods for acquisition, reconstruction, and analysis using RT-MRI. The block diagram of a system that employs such methods is illustrated in Figure 2(b). This system uses sparse sampling of k-space, motivated by the fact that even with undersampled spiral k-space data non-Cartesian parallel imaging can recover images free from aliasing artifacts.

Spatial weighting produced by multiple coil arrays can be used for additional encoding (in addition to the Fourier transform) to remove aliasing and improve temporal resolution. Alternatively, temporal filtering can be employed, given prior knowledge that certain portions of the object are static or require reduced temporal bandwidth. These approaches, which include UNFOLD, k-t BLAST, and k-t SENSE, have been applied recently to speech imaging using radial trajectories with reasonable success.

In the same context of accelerated acquisition illustrated in Figure 2(b), the reconstruction step involves



[FIG5] (a) Word initial [n]: "TOE NODE." (b) Word final [n]: "BONE OH." Typical velum aperture (top row) and tongue tip constriction (bottom row) time functions for different positions of the nasal consonant.

compressed sensing. This emerging technique has provided a theoretical framework in which a signal can be recovered by minimizing the l_1 norm of a sparse representation in a highly underdetermined linear system provided that it is compressible in certain transform domain. Hence, the MR compressed sensing can be expressed as an unconstrained optimization

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|\Phi x - y\|_{l_2}^2 + \beta \|\Psi x\|_{l_1}, \tag{4}$$

where x is a one-dimensional array vector representing an image to be estimated, Φ is spatial encoding matrix, y is acquired undersampled k-space data in vector form. Ψ is a sparsifying basis in matrix form and β is a parameter that controls the degree of influence of sparseness in the solution. This approach has been successfully applied with an acceleration factor of 5–10 to three-dimensional angiography using variable density random subsampling on Cartesian grids.

Compared to parallel imaging and temporal filtering, the performance of compressed sensing reconstruction is highly dependent on the choice of the sparsifying basis and choice of the k-space sampling pattern (as shown in Figure 3), which may be both application specific. The sampling pattern determines the characteristics of underdetermined linear system matrix and thus affects the overall performance of compressed sensing reconstruction [5]. Moreover, k-space trajectories represent the continuous integration of gradient waveforms, which are always constrained by hardware limits such as amplitude and rate of change. An appropriate selection of the sparsifying basis needs to be specific to the upper airway images. The successful application of compressed sensing reconstruction to speech MRI will require (speech) application-specific optimization of the sparsifying basis and sampling pattern (see Figure 3).

In summary, alias-free images, which are reconstructed from the state-of-the-art acquisition/reconstruction methods, can be used to detect air-tissue boundaries by using the aforementioned conventional edge detection methods. However, the process flow in Figure 2(b) now incorporates two optimization problems instead of one, one for the accelerated total variation reconstruction, and one for the contour detection.

ANALYSIS-DRIVEN ACQUISITION, RECONSTRUCTION AND EDGE DETECTION

Dynamic speech MRI for speech production research requires the data acquisition process to push the limits of spatial resolution, temporal resolution, and real-time anatomical region coverage. The conventional approach from data acquisition to vocal tract analysis, as discussed in the previous two sections and shown in Figure 2(a) and (b), involves several stages performed in tandem:

1) Reconstruct artifact-free images from the acquired data.
2) Compute the magnitude of the reconstructed images.
3) Perform edge detection and/or contour tracking.
4) Compute vocal tract parameters.

Each stage is designed and optimized in relative isolation, resulting in excessive requirements for each step, that may not be required to meet the overall goal. In other words, this conventional approach will produce high-quality intermediate data, when computing high quality final data (vocal tract parameters) is the primary goal. To address this limitation, a "holistic" approach of jointly optimizing all stages, or optimizing a reconstruction/processing technique whose input is k-space data and outputs are the final vocal tract parameters is illustrated in Figure 2(c).

Based on recent work [6] that has shown that an optimized k-space subsampling scheme can be devised that maintains accurate detection of certain air-tissue boundaries, Canny edge detection was reinterpreted in the Fourier domain by performing the convolution of images with directional gradient masks as the multiplication of the k-space data with the Fourier transform of that mask. This led to subsampling of a ring-shaped region in k-space that maximally overlapped with the region of support of the directional masks in k-space. Fourier-based edge detection in conjunction with ring-shaped subsampling allowed for a 50% reduction in data acquisition and maintained detection of edges from broad piecewise smooth structures such as the tongue.

SPEECH PRODUCTION MODELING

Upon the successful contour extraction from the RT MRI image sequences, one can proceed with the modeling of the dynamic speech production process using the extracted contours as "features." At a simple level, following the acoustic source-filter theory of speech production, the vocal tract dimensions extracted from the images can be used to estimate the resonance frequencies of the vocal tract filter corresponding to the sounds being produced, and the time evolution of the resonance frequencies can be studied. But the vocal tract contour data also allow for a variety of other modeling possibilities, such as shape-based analyses and tract variable-based models. Recent approaches include shaping based models and tract variable based models.

SHAPING BASED MODELS

In [7], the effect of emotions on the vocal tract shaping during speech production was investigated. Four different emotions, neutral, angry, sad, and happy, were simulated by a subject for a variety of target utterances. The tongue contours were extracted from the RT-MRI recordings and a principal component analysis (PCA) was carried out for the contour sample points. The modulation achieved by the strongest PCA component for the four different emotions is illustrated in Figure 4. From these figures we can observe that in particular the tongue tip undergoes an emotion-specific shaping. Such data allow for new model for emotional speech articulation.

TRACT VARIABLE BASED MODELS

Modeling studies in the articulatory phonology framework rely on a task dynamic model to represent the articulatory dynamics in terms of gestures that are goal-directed linguistic actions of the vocal tract [8]. In this case, the midsagittal vocal tract geometry is reduced to about a dozen tract variables, which represent degrees and locations of constrictions in the vocal tract as formed by individual anatomical components. Some sample tract variables are the lip aperture, the velum aperture, and the tongue tip constriction degree shown in Figure 1(b).

In [9], RT-MRI methods were used to determine the effect of syllable position on the coordination of velum and oral gestures in the articulation of nasal sounds like [n]. Consider, for example, the two utterances "TOE NODE," and "BONE OH." Each contains the nasal consonant [n] in a position between two [o] vowels. However, the nasal consonant is in the onset of a word in the first case and the coda of a word in the other. In both cases, the production of [n] requires an opening of the nasal passage by lowering the velum and a closure of the oral passage by moving the tongue onto the alveolar ridge. RT-MRI enables us to study such dynamic coordination in a direct and quantitative way.

The time evolution of the velum aperture and the tongue tip constriction degree for the word initial position of [n] are illustrated in Figure 5(a) (top and bottom rows, respectively). The distinct minimum in the velum aperture time function corresponds to the lowering of the velum for the production of the nasal consonant, and at around the same time the tongue tip creates the necessary occlusion of the oral cavity as indicated by a distinct peak in the aperture time function. Comparing this to the aperture time functions for the word final nasal position in Figure 5(b), the study showed that in general for word onset nasals the tongue gesture precedes the velum gesture, whereas for word final nasals the opposite is true, i.e., the velum opening precedes the tongue closure.

OUTLOOK

The availability of such dynamic speech imaging data opens up other exciting new possibilities for advancing scientific and technological work. Large volumes of direct speech production data afford the use of machine learning approaches to characterize patterns underlying complex encoding of intent and emotions in speech. Current research is directed towards jointly modeling the vocal tract shaping process and the speech signal, both of which are asynchronously related using advanced signal processing tools such as particle filters, hidden Markov models, and graphical models. Results from such efforts are anticipated to open new directions for the field of speech processing.

ACKNOWLEDGMENTS

This work was supported by NIH Grant R01 DC007124-01. We acknowledge the support and collaboration of the Speech Production and Knowledge group at the University of Southern California.

AUTHORS

Erik Bresch (bresch@usc.edu) and ***Yoon-Chul Kim*** (yoonckim@usc.edu) are Ph.D. students in the Ming Hsieh Department of Electrical Engineering at the University of Southern California (USC).

Krishna Nayak (knayak@sipi.usc.edu) is an assistant professor of electrical engineering and jointly in Biomedical Engineering and Medicine at USC and director of the Magnetic Resonance Engineering Laboratory.

Dani Byrd (dbyrd@usc.edu) is a professor in the Department of Linguistics and a vice dean of the College at USC, where she directs the Phonetics Laboratory.

(continued on page 132)

BEST PRACTICES

- DO plan ahead; write and revise a good outline before you start writing paragraphs.
- DO write the methods and results sections before expanding the introduction outline and writing the abstract and conclusions.
- DO use figures and illustrations, but as few as possible, so each can be large enough to be read without difficulty.
- DO focus on results, not on methods; e.g., instead of saying "We use machine learning techniques to study ways to improve A," say "We get an X % reduction in error using machine-learning technique Z."
- DO follow the composition rules in [1]–[6].
- DO spend extra time at the end to refine the introduction, abstract, and conclusions.
- DO use the active voice; it strengthens actions and thus the argument. For example, instead of saying "the input signal is processed by the preconditioning filter A," say "filter A removes high-frequency noise from the input signal;" this eliminates the redundant "is processed" and adds the reason for filtering.
- DO revise and edit; cut, cut, cut!
- DO use spelling and grammar checking.
- DO spend time to make your illustrations visually pleasing.
- DO use a direct, factual tone; breezy or opinionated descriptions are distracting.
- DO break long sections into subsections; subsection titles should be brief and draw attention to key points.
- DO use the conclusions section to emphasize the main contributions of your research; don't include a summary of what was described in the methods section.
- DO use certain terms with care; say "to optimize performance" or "to maximize usefulness" only if the system does indeed maximize an appropriate metric.
- At many stages during the writing process, take a break, "step out of the paper," and think like a reader. Reread what you wrote with a critical eye to find parts that are unclear, unjustified, or not stressed well enough.

takes time, and we should budget for it; as Blaise Pascal once said in a letter to a friend, "I have made this letter longer than usual, only because I have not had the time to make it shorter." A compilation of other good suggestions from [1]–[6], plus some other tips, are shown in the "Best Practices" sidebar.

CONCLUSIONS

The impact of a technical article usually depends as much on how well it is written as on its technical contributions. The main aspect of good technical writing is doing it for the reader: you write what you believe the reader wants to hear, not just what you want to say. Careful planning, good structuring, and concise and clear presentation of ideas and illustrations are common marks of good articles. Close attention to the points discussed here can help you improve the impact of your next article. Your readers will thank you for that.

AUTHOR

Henrique S. Malvar (malvar@microsoft.com) is a Microsoft Distinguished Engineer and the managing director of the Microsoft Research Laboratory in Redmond, Washington.

REFERENCES

[1] W. Strunk and E.B. White, *The Elements of Style,* 3rd ed. New York: McMillan, 1979.

[2] G. Blake and R.W. Bly, *The Elements of Technical Writing.* New York: McMillan, 1993.

[3] J. Zobel, *Writing for Computer Science*. New York: Springer, 1997.

[4] D.R. Morgan, "Dos and don'ts of technical writing," *IEEE Potentials,* vol. 24, pp. 22–25, Aug.–Sept. 2005.

[5] M.M. Pierson and B.L. Pierson, "Beginnings and endings: Keys to better engineering technical writing," *IEEE Trans. Prof. Commun.,* vol. 40, pp. 299–304, Dec. 1997.

[6] IEEE Author Digital Toolbox [Online]. Available: http://www.ieee.org/web/publications/authors/transjnl/index.html

[7] I. Asimov, *Gold.* New York: Harper Collins, 1995, pt. 3, "Revisions."

[8] E.R. Tufte, *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press, 1993.

[exploratory DSP] continued from page 128

Shrikanth Narayanan (shri@sipi.usc.edu) is Andrew J. Viterbi Professor of Engineering at USC where he is a professor in electrical engineering and jointly in computer science, linguistics, and psychology, and directs the Signal Analysis and Interpretation Laboratory.

REFERENCES

[1] E. Bresch, J. Nielsen, K. Nayak, and S. Narayanan, "Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans," *J. Acoust. Soc. Amer.*, vol. 120, no. 4, pp. 1791–1794, Oct. 2006.

[2] S. Narayanan, K.S. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *J. Acoust. Soc. Amer.*, vol. 115, no. 5, pp. 1771–1776, 2004.

[3] J.I. Jackson, C.H. Meyer, D.G. Nishimura, and A. Macovski, "Selection of a convolution function for Fourier inversion using gridding," *IEEE Trans. Med. Imaging*, vol. 10, no. 3, pp. 473–478, Sept. 1991.

[4] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: A survey," *Medical Image Anal.*, vol. 1, no. 2, pp. 91–108, 1996.

[5] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Mag.*, 2008, submitted for publication.

[6] Y. Kim, J. Nielsen, S. Narayanan, and K. Nayak, "Edge detection using sub-sampled k-space data: application to upper airway MRI," in *Proc. 15th Annu. Meeting Int. Society Magnetic Resononace in Medicine*, Berlin, May 2007, p. 3458.

[7] S. Lee, E. Bresch, and S. Narayanan, "An exploratory study of emotional speech production using functional data analysis techniques," in *Proc. 7th Int. Seminar Speech Production*, Ubatuba, Brazil, Dec. 2006, pp. 11–17.

[8] C.P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3-4, pp. 155–180, 1992.

[9] S. Tobin, D. Byrd, E. Bresch, and S. Narayanan, "Syllable structure effects on velum-oral coordination evaluated with real-time MRI," *J. Acoust. Soc. Amer.*, vol. 119, no. 5, p. 3302, May 2006.