Sound synchronization and motion compensated reconstruction for speech Cine MRI
Pierre-André Vuissoz, Freddy Odille, Yves Laprie, Emmanuel Vincent, Jacques Felblinger

To cite this version:
Pierre-André Vuissoz, Freddy Odille, Yves Laprie, Emmanuel Vincent, Jacques Felblinger. Sound synchronization and motion compensated reconstruction for speech Cine MRI. ISMRM 2015 Annual Meeting, May 2015, Toronto, Canada. hal-01183504

HAL Id: hal-01183504
https://inria.hal.science/hal-01183504
Submitted on 6 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Sound synchronization and motion compensated reconstruction for speech Cine MRI.

Pierre-André Vuissoz1,2, Freddy Odille1,2, Yves Laprie3,4, Emmanuel Vincent3,5, and Jacques Felblinger6,7

1. Imagerie Adaptive Diagnostique et Interventionnelle, Université de Lorraine, Nancy, France. 2. U947, INSERM, Nancy, France. 3. LORIA, Université de Lorraine, Nancy, France. 4. LORIA, CNRS, Nancy, France. 5. LORIA, INRIA, Nancy, France. 6. University Hospital Nancy, Nancy, France. 7. CIC-IT 1433, INSERM, Nancy, France

Purpose:
Dynamic imaging of the vocal tract is important for modeling speech through the acoustic-articulatory relation. The average duration of each sound is about 80 ms. Movements of each articulator, in particular the tongue, should be captured with sufficient precision.

Previous works:
- X-ray video fluoroscopy: ionizing radiation;
- Real-time MRI: limited SNR and spatial resolution;
- Cine MRI with acoustic device gating: needs highly reproducible motion.

This work:
- Ungated acquisition with acoustic device recording;
- Motion-compensated cine reconstruction.

Material & Methods:
Data acquisition:
Dynamic MRI data were acquired at 3T (Signa HDxt, GE Healthcare, Milwaukee, WI) using an ungated balanced SSFP sequence (one sagittal slice, 256x256 matrix, TR/TE = 3.8/1.7 ms, 5 mm slice thickness, 45° flip angle, FOV 30 cm, 65 temporal phases).

A list of 10 sentences was carefully selected for the dynamic imaging protocol to be short and yet yield a good coverage of the tongue movements in French language:
- “Ma chemise est roussie”,
- “Donne un petit coup”,
- “Voilà des bougies”,
- “Une réponse ambiguë”,
- “Louis pense à ça”,
- “Mets tes beaux habits”,
- “Une pâte à choux”,
- “Prête-lui seize écus”,
- “Chevalier du gué”,
- “Il fume son tabac”.

For each dynamic acquisition the subject was asked to repeat one of the 10 sentences until the sequence stopped. Acoustic signals were acquired using an optical microphone (FOMRI III, Optoacoustics, Yehuda Israel). The scanner’s acquisition window signal was also recorded with the device to allow synchronization of MR events and acoustic signals.

Acoustic signal processing:
Microphone recordings were first denoised to eliminate gradient noise, using a general audio source separation framework with incorporation of prior knowledge. Then they were phonetically segmented by hand in order to annotate the beginning of each phone within the sound record. An acoustic phase signal was then created in order to indicate the temporal position within the sentence.

In order to account for the variability of repeated phone timings, the phase signal was adjusted using a piecewise linear scaling based on the manual segmentation.

Reconstruction:
Cine images from each template sentence were reconstructed using cine-GRICS [7]. Here cine-GRICS used the acoustic phase signal to reconstruct the images by a motion-compensated sliding-window approach. The window width was 80 ms and motion correlated signals were the relative phase distance to the key frame position and the squared distance. 128 key frames were used.

Acoustic signal processing

Fig 1.: Two temporal positions (A) 1 and (B) 43 of the 128 image cine loop of sagittal slice along the vocal tract with TM positions.

Results:
- In Figure 1 two images of the subject pronouncing “Voilà des bougies” show the absence of motion blurring or artifact in the reconstructed cine loop.
- In Figure 2 the distances between the tongue dorsum and the hard palate and between the tongue back and the pharyngeal wall are shown along with the template sentence signal.

Discussion and conclusion:
- A potential issue with balanced-SSFP is related to the banding artifacts and the strong B0 gradient at the air tissue interface especially in the tongue, but the signal void in the vocal tract is still clearly distinguished in the dynamic movie.
- Each cine loop enables the delineation of the vocal tract with sufficient spatial and temporal resolution enabling the acquisition of a personalized speech model within an MR examination of half an hour.

References:
[3] Odille et al. MRM. 60:146-157 (2008);
[4] Batchelor et al. MRM. 54:1273-1280 (2005);