

Real-time speech MRI: what is the optimal temporal resolution for clinical velopharyngeal closure assessment?

Matthieu Ruthven^{1,2}, Andreia C. Freitas³, Stephen F. Keevil^{2,4}, and Marc E. Miquel¹

¹*Clinical Physics Department, Barts Health NHS Trust, London, United Kingdom,* ²*Imaging Sciences & Biomedical Engineering Research Division, King's College London, London, United Kingdom,* ³*William Harvey Research Institute, Queen Mary University of London, London, United Kingdom,* ⁴*Medical Physics Department, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom*

Synopsis

Clinical velopharyngeal closure assessment involves imaging patients while they perform standard speech tasks. Real-time MRI could offer an alternative to the imaging techniques used at present, however, there is currently no consensus on the optimal temporal resolution. The purpose of this study is to determine an optimal temporal resolution by comparing the numbers of velopharyngeal closures in high temporal resolution and simulated lower temporal resolution datasets of healthy adult volunteers. The results of this study suggest that the optimal temporal resolution is between 7.5 and 10 frames per second. Future work will aim to pinpoint and validate this resolution.

Purpose

Velopharyngeal closure, which is when the velum and pharyngeal walls come into contact and block the opening between the oral and nasal cavities, is required for comprehensible speech. Clinical assessment of velopharyngeal closure involves imaging patients while they perform standard speech tasks. In the UK, imaging is most commonly performed using x-ray videofluoroscopy at a temporal resolution of 15 frames per second (fps) [1]. The results of studies suggest that imaging could be performed using real-time MRI (rt-MRI) [2-6], however, as highlighted in a recent article giving recommendations for real-time speech MRI [7], there is currently no consensus on the minimum temporal resolution required to capture the velopharyngeal closures that should occur in speech tasks. Determining this resolution would be beneficial for two reasons. Firstly, imaging at or higher than this resolution is required to prevent misdiagnoses due to insufficient imaging rates. Secondly, because of the trade-off between the acquisition of spatial and temporal information in MRI, imaging at resolutions as close as possible to the minimum could enable the acquisition of additional spatial data that could provide extra clinically relevant information. This temporal resolution could therefore be considered as the optimal resolution. The purpose of this study is to determine an optimal temporal resolution for clinical velopharyngeal closure assessment.

Methods

91 rt-MRI datasets of healthy adult volunteers (age range 24 to 50 years) acquired at Barts Health NHS Trust for ethics committee approved speech studies between 2010 and 2015 were identified and retrospectively analysed. 40 additional datasets have also been identified but not yet analysed. All datasets consist of 10mm thick mid-sagittal slices acquired while volunteers performed speech tasks. Images were acquired using either a 1.5T Achieva (67 datasets) or a 3T TX Achieva (24 datasets) scanner (Philips Healthcare, Best, the Netherlands) in conjunction with a 16-channel neurovascular coil. Balanced steady state free precession pulse sequences were used at 1.5T and fast low-angle shot pulse sequences at 3T.

34 datasets have a temporal resolution of 10fps, 26 a resolution of 15fps and 31 a resolution of 20fps. For all 91 datasets, the speech task included counting from one to ten and phonating nonsense (“za-na-za”, “zu-nu-zu”, “zi-ni-zi”) with volunteers instructed to speak at a normal rate. For a subset of 24, the speech task also included standard clinical velopharyngeal closure assessment sentences (“Bob is a baby boy”, “I saw Sam sitting on the bus”, “Tim is putting a hat on”) and volunteers also repeated the whole speech task at a faster rate.

N/2 and N/3fps datasets (where N=10, 15 or 20) were simulated from each dataset using three methods implemented using Matlab (R2015a, MathWorks, Natick, MA): decimation, averaging and block filling (see figures 1 and 2). The numbers of closures in original and simulated datasets were determined from intensity-time plots (see figure 3) generated using Matlab. The numbers of closures in original datasets acquired at different temporal resolutions were compared using one-way analyses of variance. SPSS (v22, IBM, Armonk, NY) was used for all statistical analyses. Groups of datasets simulated from the same original dataset were analysed for closure losses using the method shown in figure 4.

Results

For the counting and the nonsense phonation, there were no statistically significant differences between the mean numbers of closures in the original 10, 15 and 20fps datasets.

It is worth noting that, regardless of acquisition rate, natural variability in speech can cause both intra- and inter-volunteer variations in the number of closures.

There were closure losses in 27 of the 91 groups of simulated datasets. The temporal resolutions at which these losses first occurred, and the section(s) of the speech tasks in which these occurred are shown in figure 5. The highest temporal resolution at which closure losses first occurred was 7.5fps.

Discussion and conclusion

So far, the results of this study suggest that the optimal temporal resolution is between 7.5 and 10fps for counting, for nonsense phonation, and for the test sentences. This is slightly higher than what could be inferred from a previous study comparing rt-MRI at 6fps and clinical x-ray videofluoroscopy (imaging rate not given) [6], but is in the range suggested by [7]. Once the 40 additional datasets have been analysed, we aim to pinpoint the optimal temporal resolution and validate it by comparing the numbers of closures in new datasets acquired at the optimal temporal resolution and at slightly higher and lower temporal resolutions.

References

- [1] Sell D, Pereira V (2011) ‘Instrumentation in the analysis of the structure and function of the velopharyngeal mechanism’ in *Cleft palate speech: assessment and intervention* Wiley, Chichester
- [2] Scott AD, Boubertakh R, Birch MJ, Miquel ME (2012) Towards clinical assessment of velopharyngeal closure using MRI: evaluation of real-time MRI sequences at 1.5 and 3T, *British Journal of Radiology* **85**:e1083-e1092.

[3] Silver AL, Nimkin K, Ashland JE, Ghosh SS, van der Kouwe AJW, Brigger MT, Hartnick CJ (2011) Cine magnetic resonance imaging with simultaneous audio to evaluate pediatric velopharyngeal insufficiency. *Archives of Otolaryngology Head & Neck Surgery* **137**:258-263

[4] Maturo S, Silver A, Nimkin K, Sagar P, Ashland J, van der Kouwe AJW, Hartnick C (2012) MRI with synchronized audio to evaluate velopharyngeal insufficiency. *Cleft Palate Craniofacial Journal* **49**:761-763.

[5] Drissi C, Mitrofanoff M, Talandier C, Falip C, Le Couls V, Adamsbaum C (2011) Feasibility of dynamic MRI for evaluating velopharyngeal insufficiency in children. *European Radiology* **21**:1462-1469.

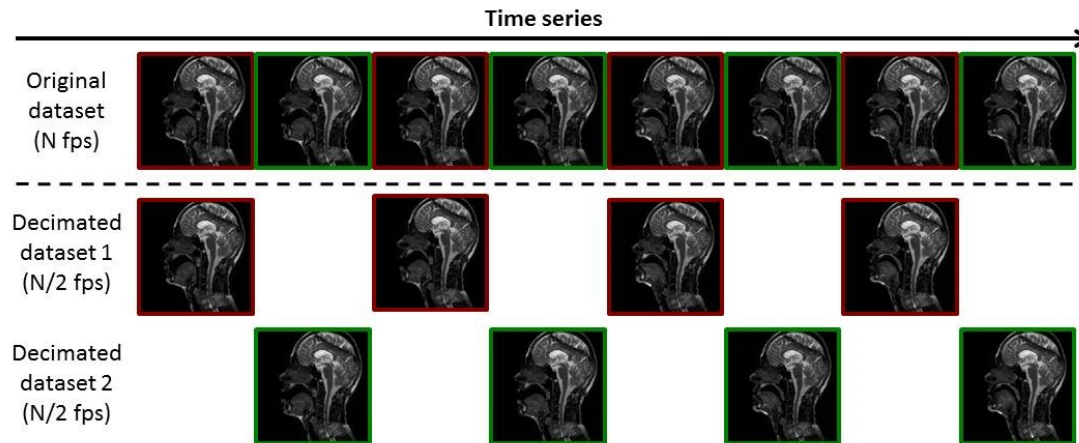
[6] Beer AJ, Hellerhoff P, Zimmermann A, Mady K, Sader R, Rummeny EJ, Hannig C (2004) Dynamic near-real-time magnetic resonance imaging for analysing the velopharyngeal closure in comparison with videofluoroscopy. *Journal of Magnetic Resonance Imaging* **20**:791-797.

[7] Lingala SG, Sutton BP, Miquel ME, Nayak KS (2016) Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging* **43** (1): 28-44, DOI: 10.1002/jmri.24997

Fig 1. A diagram showing how $N/2$ fps datasets were simulated from N fps datasets using image space methods. $N/3$ fps datasets were simulated in similar ways.

Image space simulation methods

A) Decimation:



B) Averaging:

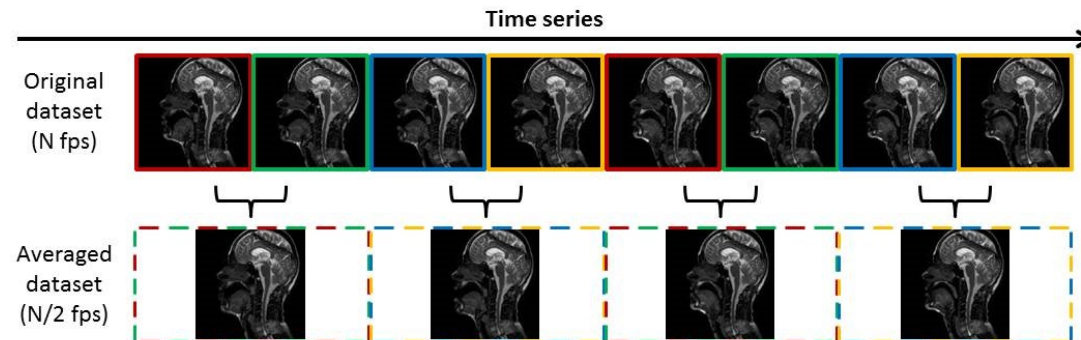


Fig 2. A diagram showing how N/2 and N/3fps datasets were simulated from Nfps datasets using a k-space method

K-space simulation method

Block filling:

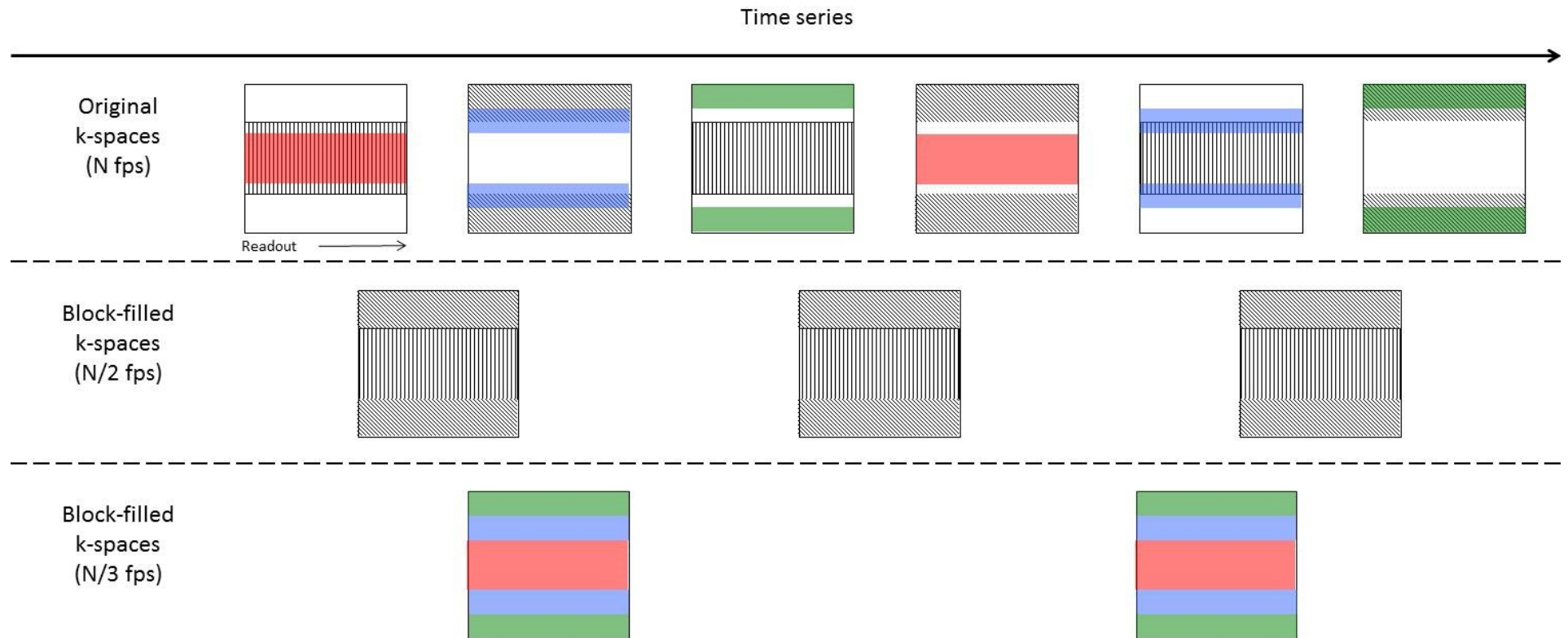
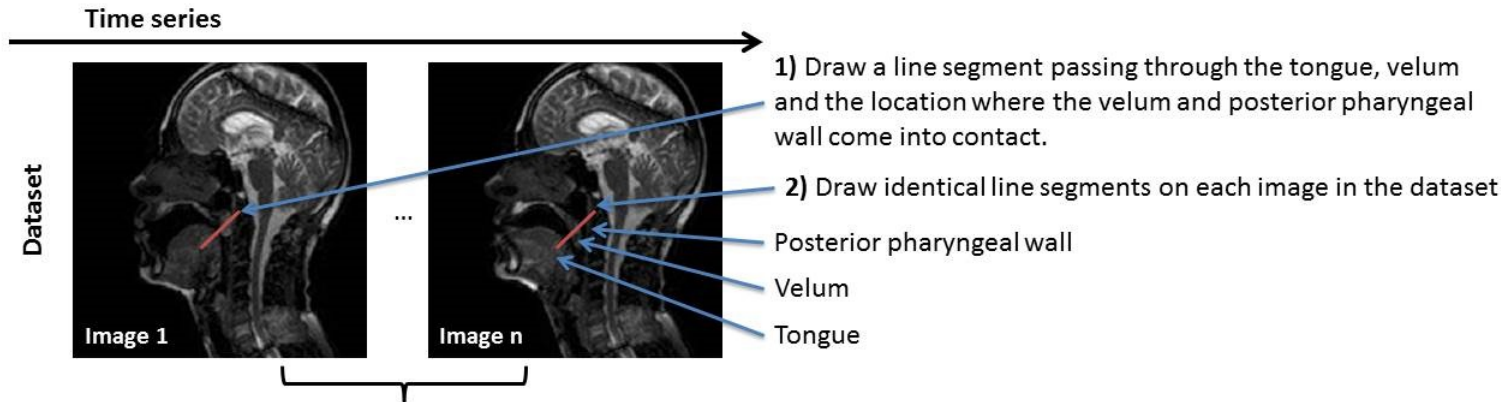


Fig 3. A diagram showing how intensity-time plots were generated from datasets

Method used to generate intensity-time plots



3) Stack the line segments to generate an intensity-time plot:

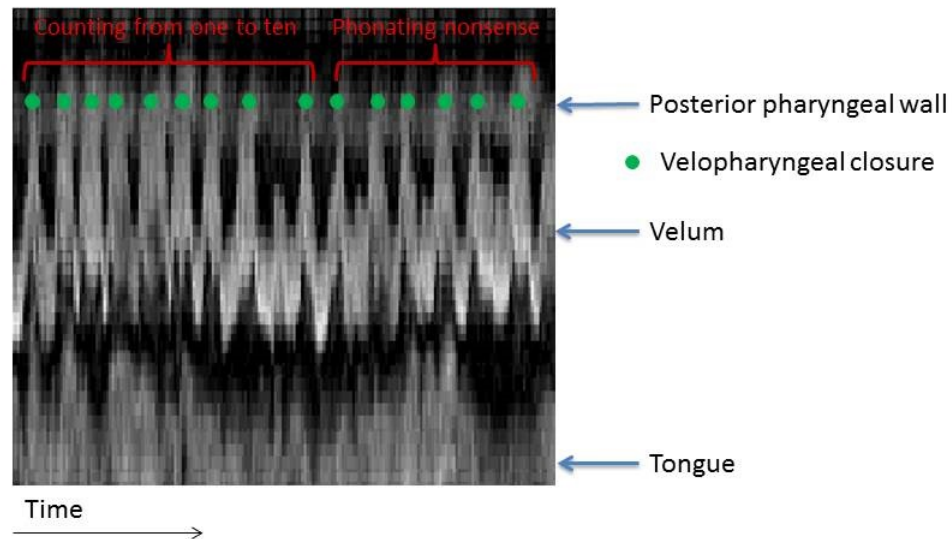


Fig. 4 A flow chart showing how groups of datasets simulated from the same original dataset were analysed in order to identify losses in velopharyngeal closures

Method used to analyse each simulated dataset group

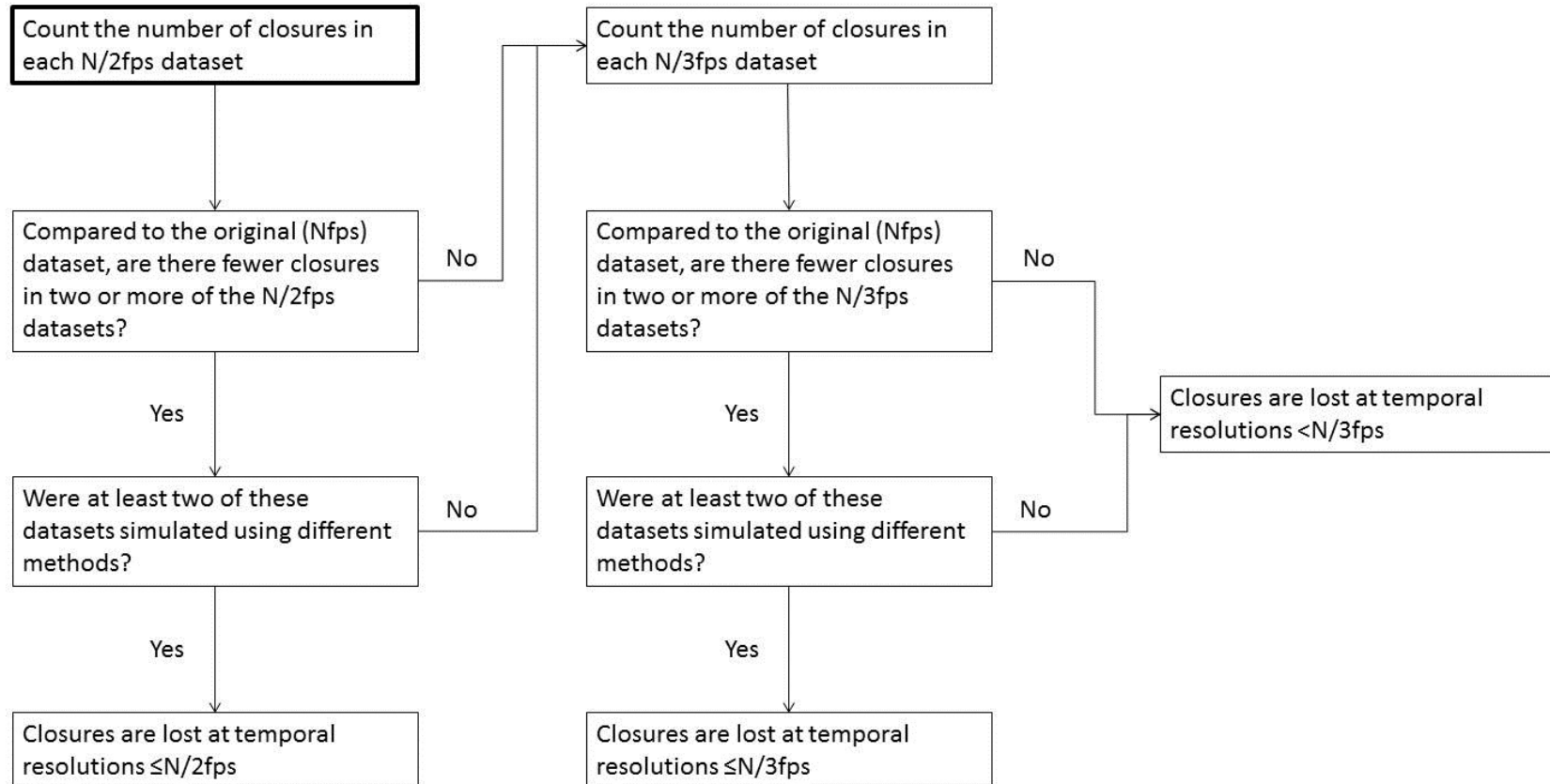


Fig 5. A bar chart showing the temporal resolutions at which closure losses first occurred, and the section(s) of the speech tasks in which these occurred

