Radiotherapy and Oncology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

The impact of audiovisual biofeedback on 4D functional and anatomic imaging: Results of a lung cancer pilot study

Jaewon Yang [a,c,1], Tokihiro Yamamoto [d,1], Sean Pollock [e], Jonathan Berger [b], Maximilian Diehn [c], Edward E. Graves [c], Billy W. Loo Jr. [c], Paul J. Keall [e,*]

[a] Department of Electrical Engineering; [b] Center for Computer Research in Music and Acoustics, Stanford University; [c] Department of Radiation Oncology, Stanford University School of Medicine; [d] Department of Radiation Oncology, University of California Davis School of Medicine, Sacramento, USA; and [e] Radiation Physics Laboratory, University of Sydney, Australia

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 29 March 2016
Accepted 11 May 2016
Available online xxxx

Keywords:
Motion management
Audiovisual biofeedback
Lung cancer
4D PET
4D CT

ABSTRACT

Background and purpose: The impact of audiovisual (AV) biofeedback on four dimensional (4D) positron emission tomography (PET) and 4D computed tomography (CT) image quality was investigated in a prospective clinical trial (NCT01172041).

Material and methods: 4D-PET and 4D-CT images of ten lung cancer patients were acquired with AV biofeedback (AV) and free breathing (FB). The 4D-PET images were analyzed for motion artifacts by comparing 4D to 3D PET for gross tumor volumes (GTV_{PET}) and maximum standardized uptake values (SUV_{max}). The 4D-CT images were analyzed for artifacts by comparing normalized cross correlation-based scores (NCCS) and quantifying a visual assessment score (VAS). A Wilcoxon signed-ranks test was used for statistical testing.

Results: The impact of AV biofeedback varied widely. Overall, the 3D to 4D decrease of GTV_{PET} was 1.2 ± 1.3 cm^3 with AV and 0.6 ± 1.8 cm^3 for FB. The 4D-PET increase of SUV_{max} was 1.3 ± 0.9 with AV and 1.3 ± 0.8 for FB. The 4D-CT NCCS were 0.65 ± 0.27 with AV and 0.60 ± 0.32 for FB ($p = 0.08$). The 4D-CT VAS was 0.0 ± 2.7.

Conclusion: This study demonstrated a high patient dependence on the use of AV biofeedback to reduce motion artifacts in 4D imaging. None of the hypotheses tested were statistically significant. Future development of AV biofeedback will focus on optimizing the human–computer interface and including patient training sessions for improved comprehension and compliance.

Respiratory motion causes artifacts in positron emission tomography (PET) and computed tomography (CT). In PET imaging, respiratory motion artifacts may result in underestimation of the corresponding standardized uptake value (SUV) and overestimation of the tumor volume [1]. In CT imaging, respiration may cause the four types of artifacts (blurring, duplicate structure, overlapping structure, and incomplete structure) which potentially deteriorate the quality of CT images [2]. Thus, respiratory motion management is needed to mitigate such artifacts, which is achieved through four-dimensional (4D) imaging.

However, 4D imaging may fail to reduce artifacts if breathing is irregular during data acquisition due to inadequate respiratory motion sampling for image reconstruction [2,3]. For this reason, respiratory training or coaching is a commonly used technique to reduce irregularity in breathing cycles. Among the techniques, audiovisual (AV) biofeedback is a real-time, interactive and personalized system designed to help a patient self-regulate their breathing, demonstrating a reduction in the average cycle-to-cycle variations in respiratory amplitude and period by up to 50% and 70%, respectively [4]. To date, only two studies have investigated the impact of breathing training on anatomic imaging such as MRI and CT [5,6]; however, no investigation has been performed for the impact of AV biofeedback on functional imaging. Therefore, for the first time, the impact of AV biofeedback was evaluated on 4D-PET and 4D-CT imaging in a patient study.

Materials and methods

Data acquisition

In an institutional review board (IRB)-approved prospective clinical trial (NCT01172041), eligibility criteria for 10 lung cancer patients (age $\geqslant$ 18) included a diagnosis of AJCC Stage I–IV lung cancer of any histology and Karnofsky Performance Status $\geqslant$ 50. The patients were scanned in pre-treatment, except patients 4

* Corresponding author at: Radiation Physics Laboratory, Sydney Medical School, The University of Sydney, Camperdown, NSW 2006, Australia.
E-mail address: paul.keall@sydney.edu.au (P.J. Keall).
[1] Equally contributing first authors.

http://dx.doi.org/10.1016/j.radonc.2016.05.016
0167-8140/

Table 1
Scan time of PET and scan order of 4D PET and 4D CT information for the 10 lung cancer patients recruited to the audiovisual (AV) biofeedback 4D imaging study.

Patient #	PET scan duration (# of bed positions)		PET scan start time difference	Scan order
	AV	FB		
1	18 min (3)	30 min (6)	24.9 min	$FB_{PET} \rightarrow AV_{PET}FB_{PET} \rightarrow FB_{CT}FB_{PET} \rightarrow AV_{CT}$
2	15 min (3)	16.5 min (3)	18.5 min	$AV_{PET} \rightarrow FB_{PET} \rightarrow AV_{CT} \rightarrow FB_{CT}$
3	22.5 min (3)	21 min (3)	26 min	$FB_{PET} \rightarrow AV_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT}$
4	11 min (2)	10 min (2)	8.3 min	$AV_{PET} \rightarrow FB_{PET} \rightarrow AV_{CT} \rightarrow FB_{CT}$
5	11 min (2)	10 min (2)	22.2 min	$FB_{PET} \rightarrow AV_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT}$
6	16.5 (3)	15 min (3)	22.5 min	$FB_{PET} \rightarrow AV_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT}$
7	7 min (1)	21 min (3)	36.4 min	$FB_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT} \rightarrow AV_{PET}$
8	6 min (1)	15 min (3)	44 min	$FB_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT} \rightarrow AV_{PET}$
9	5 min (1)	12 min (3)	43.5 min	$FB_{PET} \rightarrow FB_{CT} \rightarrow AV_{CT} \rightarrow AV_{PET}$
10	6 min (1)	12 min (3)	12.9 min	$FB_{PET} \rightarrow AV_{PET} \rightarrow AV_{CT} \rightarrow FB_{CT}$

(post-treatment) and 5 (mid-treatment), with arms raised above their head on a Discovery ST PET/CT scanner (GE Healthcare, Waukesha, WI) along with a real-time position management (RPM) system (Varian Medical Systems, Palo Alto, CA). In each scan, 4D-PET and 4D-CT data with AV biofeedback (AV) and free breathing (FB) were acquired consecutively in the same session [4].

4D-PET and 4D-CT imaging

In 4D PET, PET raw data were acquired in list-mode for respiratory-correlated gating [7]. Data acquisition time for the second 4D-PET scan was increased to compensate for total activity decaying (Table 1). The 4D-PET data were sorted into six bins by phase-based sorting using GE Respiratory Gating Toolbox. For attenuation correction (AC), averaged-CT (ACT) was consistently employed to reduce a spatial misalignment, caused by irregular respiratory motion, between PET and CT. The exceptional use of six-bin gated CT was allowed for patient 8 because ACT failed to cover the tumor motion trajectory (>20 mm). PET images were reconstructed through the Ordered Subset Expectation Maximization (OSEM) algorithm (21 subsets, 2 iterations, and 6 mm FWHM Gaussian post filter) with the voxel size of $2.3 \times 2.3 \times 3.75$ mm^3. Additionally, 4D CT were acquired in cine mode with parameters as follows: 120 kVp, approximately 100 mAs per slice, 0.5 s gantry rotation, 0.45 s cine interval, and 8 slices with the voxel dimension of $1.0 \times 1.0 \times 2.5$ mm^3 [8]. The oversampled CT slices were then sorted into ten respiratory bins by phase-based sorting [9].

Quantification

In 4D-PET analysis, hypotheses are (H1) AV biofeedback increases GTV difference between 3D and 4D PET more than FB and (H2) AV biofeedback increases SUV_{max} difference between 3D and 4D PET more than FB, based on the commonly accepted ideas that 4D-PET GTV is smaller than 3D-PET GTV and that 4D-PET SUV_{max} is higher than 3D-PET SUV_{max} [10]; these differences are expected to be increased if respiratory motion becomes more regular. In the analysis, only tumors showing motion greater than 5 mm, the suggested level at which explicit motion management should be considered, were included as PET signals were focused in the tumor area [11]. Although there is no standardized method for automatic segmentation of GTV from PET images (GTV_{PET}) due to the low spatial resolution and high noise characteristics of PET images [12], region growing was employed as a compromise of simplicity and complexity among various approaches [13]. To derive the GTV_{PET}, each tumor volume was segmented automatically by region growing with various percentage thresholds (10–90%, steps of 10%) of SUV_{max}, and GTV_{PET} closest to GTV_{CT} was considered as an optimal tumor volume. This method was applied consistently for all patients except patient 3 in which the region growing algorithm did not result in a closed surface for the free breathing images, yielding an undefined volume. To allow the inclusion of patient 8's data and ensure consistency between AV and FB in the study, a higher threshold was used for both AV and FB, of 80%. This necessarily yielded a smaller GTV_{PET} than GTV_{CT} for this patient.

In 4D-CT analysis, hypotheses are (H3) AV biofeedback reduces artifacts determined via normalized cross correlation-based score (NCCS) and (H4) AV biofeedback reduces artifacts determined via visual assessment-based score (VAS). NCCS and VAS evaluated artifacts at the 0% (peak-inhale), 30% (mid-exhale), 50% (peak-exhale) and 80% (mid-inhale) phases. A NCC-based metric has been demonstrated to replicate the findings of human observers [14]. The VAS was determined by an observer (medical physicist) through comparing five pairs of coronal CT slices displayed side by side and blinded to the breathing method. The observer marked the slice that appeared to have fewer artifacts, yielding a score that represents how many times AV was selected to have fewer artifacts than FB and ranges from −5 to +5; i.e. positive scores mean AV has fewer artifacts than FB. All 10 patients were included in the CT analysis.

Respiratory regularity, consistency of breathing patterns, was quantified through the root mean square error (RMSE) of

Table 2
AV biofeedback (AV) vs. free breathing (FB) for percentage decrease in GTV_{PET}, percentage increase in SUV_{max}, RMSE displacement, and RMSE period. Percentage is quantified from the ratio of 4D-PET-measured value to 3D-PET-measured value.

Patient #	Decrease of tumor volume (%)		Increase of SUV_{max} (%)		RMSE displacement (mm)		RMSE period (sec)	
	AV	FB	AV	FB	AV	FB	AV	FB
2	22 ± 15%	15 ± 12%	18 ± 8.7%	13 ± 8.6%	0.6	1.1	0.4	0.9
6	9.4 ± 5.8%	16 ± 4.7%	9.5 ± 3.9%	14 ± 3.7%	1.7	1.2	1.9	2.0
8	64 ± 11%	75 ± 11%	26 ± 24%	35 ± 12%	1.1	3.0	0.6	1.3
9	−28 ± 187%	27 ± 67%	24 ± 19%	25 ± 11%	2.7	1.5	1.5	1.1
10	47 ± 24%	−36 ± 112%	20 ± 12%	8.9 ± 8.6%	2.6	2.4	1.6	4.3

Please cite this article in press as: Yang J et al. The impact of audiovisual biofeedback on 4D functional and anatomic imaging: Results of a lung cancer pilot study. Radiother Oncol (2016), http://dx.doi.org/10.1016/j.radonc.2016.05.016

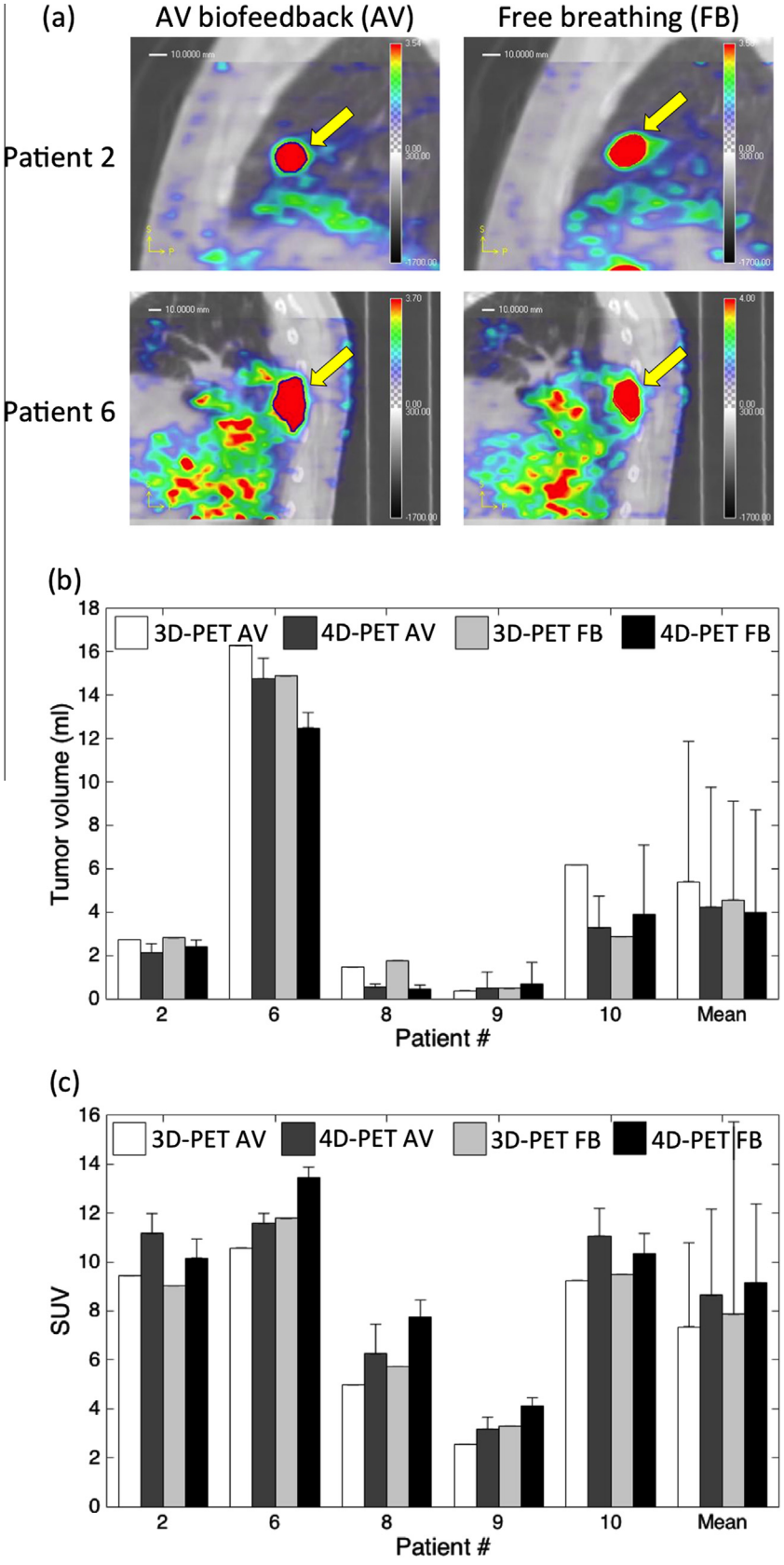

Fig. 1. AV biofeedback (AV) vs. free breathing (FB) for 4D-PET. (a) Comparison of the AV and FB 4D-PET tumor images at peak inhalation (bin-6) for patients 2 and 6; (b) Comparison of $GTV_{3D\text{-}PET}$ and $GTV_{4D\text{-}PET}$; (c) Comparison of $SUV_{max,3D\text{-}PET}$ and $SUV_{max,4D\text{-}PET}$. The values for 4D-PET were averaged across all phases for 5 patients with mobile tumors. The error bar is standard deviation.

Please cite this article in press as: Yang J et al. The impact of audiovisual biofeedback on 4D functional and anatomic imaging: Results of a lung cancer pilot study. Radiother Oncol (2016), http://dx.doi.org/10.1016/j.radonc.2016.05.016

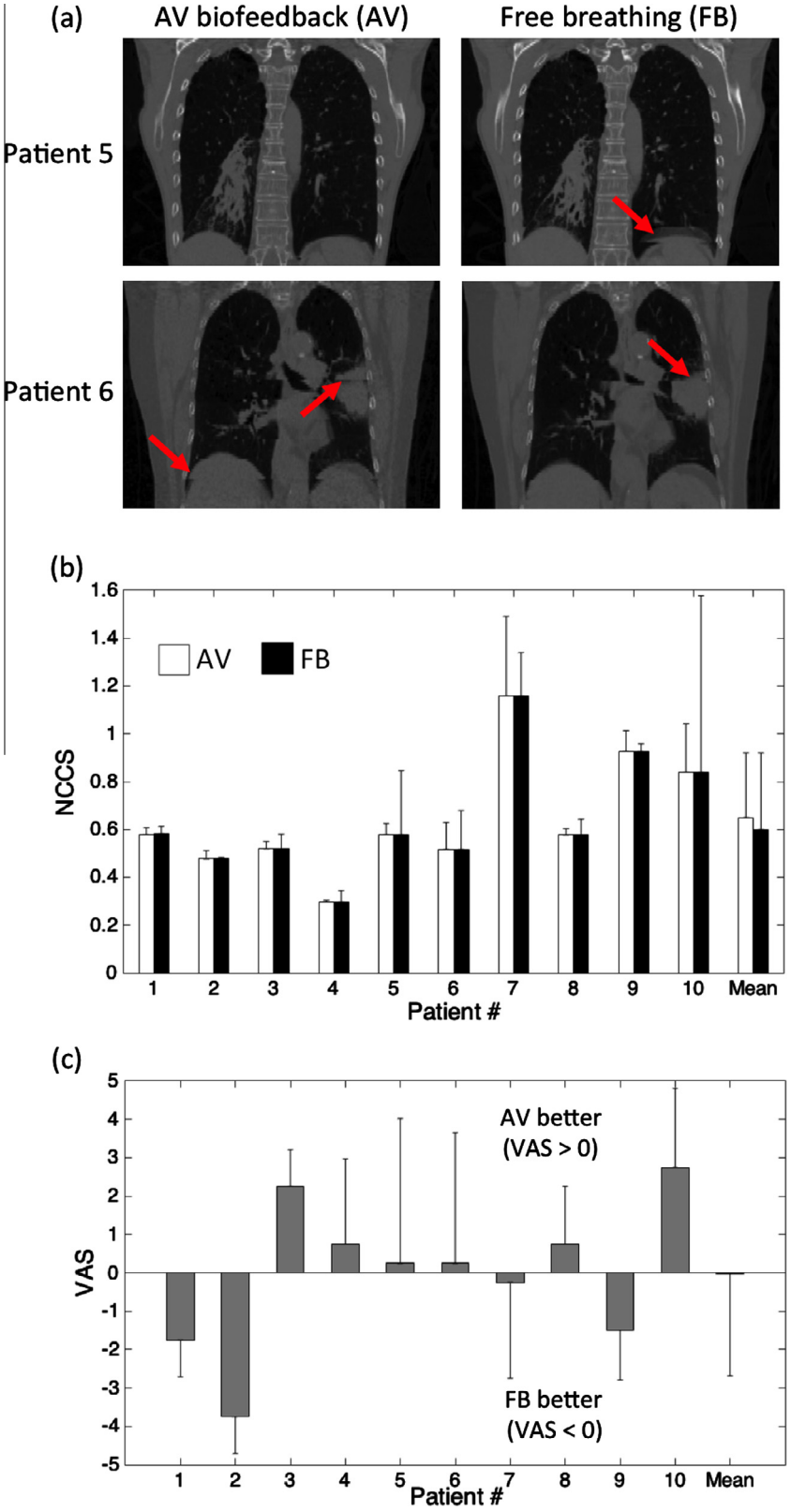

Fig. 2. AV biofeedback (AV) vs. free breathing (FB) for 4D-CT. (a) Comparison of the AV and FB 4D-CT images at peak inhalation (0% phase) for patients 5 and 6. Artifacts are denoted by red arrows; (b) Average NCCS; (c) Average VAS across all phases for 10 patients. The error bar is standard deviation.

displacement and period [4,5]. The RMSE in period was computed from the each period of the individual breathing cycles. The RMSE in displacement was calculated from displacement variation of each sample (in a breathing cycle) with respect to a corresponding sample from the post priori guiding waveform in the phase domain.

Please cite this article in press as: Yang J et al. The impact of audiovisual biofeedback on 4D functional and anatomic imaging: Results of a lung cancer pilot study. Radiother Oncol (2016), http://dx.doi.org/10.1016/j.radonc.2016.05.016

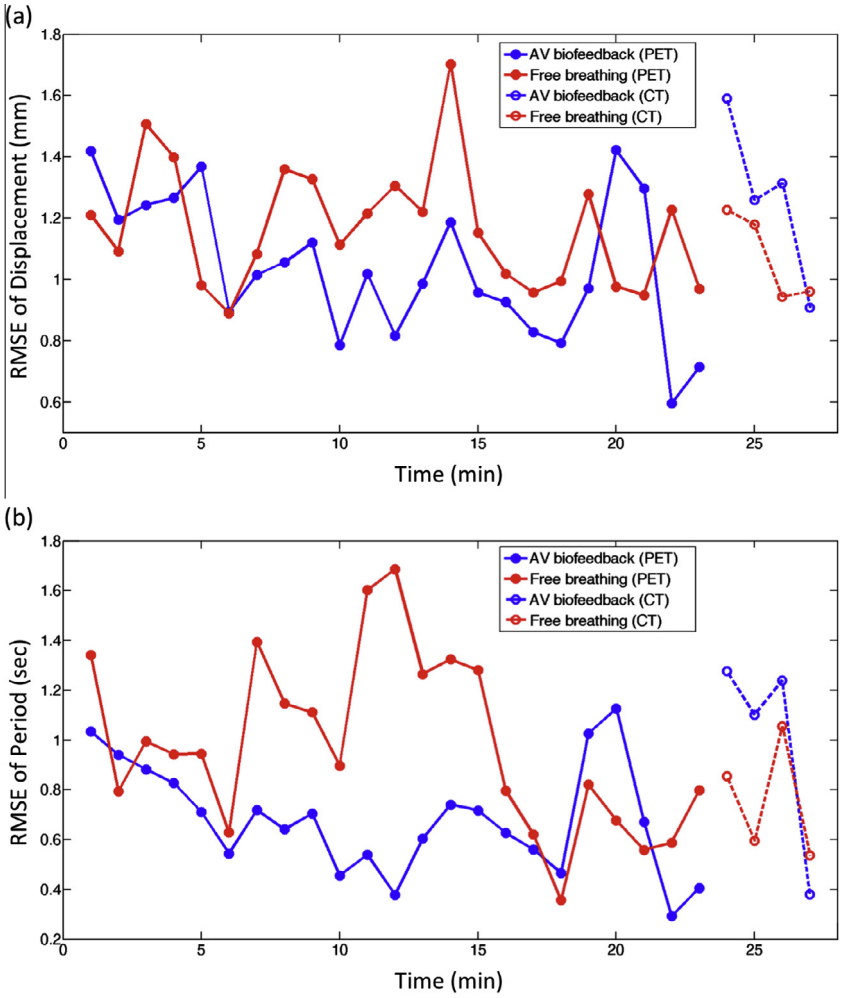

Fig. 3. AV biofeedback (AV) vs. free breathing (FB) traces during CT and PET scans: overall time-averaged RMSE of (a) displacement and (b) period for all respiratory traces with one-minute increments for 10 patients.

$$\text{RMSE in displacement} = \frac{\sum_{Allcycles}\sqrt{\sum_{i=1\ldots360}\frac{(x_i-y_i)^2}{360}}}{Total\ Cycles}$$

where x_i are the samples from each cycle of the patients waveform, y_i are the samples from the post priori calculated average waveform at phase i for every degree of phase (hence 360°). A low value of RMSE is an indicative of a highly reproducible respiratory signal.

The four hypotheses were tested statistically through a Wilcoxon signed rank test for paired samples: $p < 0.05$ indicates a statistically significant result.

Results

In 4D PET, the overall decrease of AV vs. FB GTV_{PET} (H1) was 1.2 ± 1.3 vs. 0.6 ± 1.8 cm^3, respectively, and the overall increase of AV vs. FB SUV_{max} (H2) was 1.3 ± 0.9 vs. 1.3 ± 0.8, respectively. In 4D CT, the overall result of AV vs. FB NCCS (H3) was 0.65 ± 0.27 vs. 0.60 ± 0.32, respectively ($p = 0.08$), and the overall VAS (H4) was 0.0 ± 2.7. Although all the hypotheses were not statistically significant, the results had large variations between patients (Table 2), as shown in Figs. 1 and 2.

Fig. 3 shows the results of average breathing regularity through the RMSE with one-minute increments. For 4D-PET, there was an improvement of breathing regularity with AV during first 10 min; however, the regularity decreased as training time advanced. The RMSE of 4D-PET respiratory traces (AV vs. FB) was 1.5 ± 0.7 vs. 1.6 ± 0.6 mm and 1.0 ± 0.5 vs. 1.7 ± 1.1 s ($p = 0.01$) for displacement and period, respectively. However, the RMSE of 4D-CT respiratory traces (AV vs. FB) was 1.4 ± 0.7 vs. 1.3 ± 0.6 mm and 1.0 ± 0.6 vs. 0.9 ± 0.4 s for displacement and period, respectively. The results of breathing regularity varied widely with patients such as the results of images.

Discussion

This study investigated the impact of audiovisual (AV) biofeedback on 4D-PET image quality for the first time, on a cohort of ten lung cancer patients (five with tumor motion >5 mm); 4D-CT image quality was also investigated. The results are equivocal: although a statistically significant reduction of motion blurring artifacts of AV over FB was not demonstrated for two functional 4D-PET imaging metrics, the result of patient 2 demonstrated the reduced motion-blurring artifacts in Fig. 1(a). Similarly, although a statistically significant difference between AV and FB was not demonstrated for two anatomic 4D-CT imaging metrics, the result of patient 5 demonstrated that reduced motion artifacts in Fig. 2 (a). These study results fall between the previous positive and negative studies: the 15 volunteer MRI study demonstrated significant reductions in the variability of external and internal (diaphragm) motion with AV biofeedback [5]; however, the 13 patient 4D-CT study showed no improved match of target delineation using

Please cite this article in press as: Yang J et al. The impact of audiovisual biofeedback on 4D functional and anatomic imaging: Results of a lung cancer pilot study. Radiother Oncol (2016), http://dx.doi.org/10.1016/j.radonc.2016.05.016

maximum intensity projection using breathing coaching with an abdominal or spirometer motion signal [6]. The presence of some negative results is of particular interest as it stimulates future investigation into why the intervention did not show a major improvement over no intervention.

Particularly, it was observed that the underlying free breathing respiratory signal was often so poor, which brings follow-up questions: was this simply that the patients could not comply with the AV biofeedback instructions due to limited lung function, or did they not comprehend the task that they were asked to perform? A comprehensive technology assessment of the patient experience could form part of future studies to more clearly differentiate the cause of the limited improvement [15,16]. Such a patient experience of technology study was conducted by Brédart et al. for respiratory gating [17].

Additionally, the overall poor improvement of breathing patterns with AV is not consistent with the positive results of the early version of AV biofeedback [18]. The most distinguishable discrepancies between the two studies are the average length of scan duration (19 vs. 4 min) and number of study sessions (1 vs. 5). For the lengthy scan duration, Fig. 3 shows that there is a window where the AV training appears most regular after an initial training period, and before fatigue sets in. For the number of study sessions, this study did not provide training sessions; while, the previous study involved five sessions per patient, showing the improved regularity over time per session [19]. These discrepancies demonstrate the importance of patient familiarity and optimal session length with AV biofeedback; in addition, improvements to the patient comprehension and human–computer interactions are needed to maximize the engagement and results that patients have with the AV biofeedback system.

Generally, it is accepted that the introduction of coaching or training changes the patterns of free breathing, which was demonstrated by Persson et al. [20]; while, Goossens et al. demonstrated that audio/visual coaching achieved high internal/external correlation and reproducibility, suggesting real-time visual feedback with audio coaching for breathing reproducibility [21]. Considering the contrary results from the two literatures, it is too early to conclude anything about the effect of scan ordering (e.g., FB → AV or AV → FB), on breathing patterns. However, to avoid the problems that arise where the use of AV biases the patterns of FB, scans with AV were performed after those with FB for 8 patients, except patients 2 and 4 having the AV prior to the FB (Table 1).

In conclusion, this study demonstrated a high patient dependence on the use of AV biofeedback to reduce motion artifacts in 4D imaging. None of the hypotheses tested were statistically significant. Future development of AV biofeedback will focus on optimizing the human–computer interface and including patient training sessions for improved comprehension and compliance.

Conflict of interest statement

Jaewon Yang – Support for this study was from Kwanjeong Educational Foundation (KEF), NIH/NCI R01 CA 093626, Stanford Bio-X travel grant. GE Healthcare provided a loan and support of the Respiratory Gating Toolbox (RGT) software for 4D-PET image reconstruction.

Tokihiro Yamamoto – reports grants from Free to Breathe, personal fees from Varian Medical Systems, non-financial support from Philips Medical Systems, outside the submitted work

Paul J. Keall – reports grants from US NIH/NCI R01 grant, other from GE Healthcare, grants from NHMRC Australia Fellowship, during the conduct of the study; In addition, Dr. Keall has a patent US patent 7,955,270 issued and is a Director of Respiratory Innovations.

The other authors – nothing to disclose.

Acknowledgments

Supported by NIH/NCI R01 CA 093626, Stanford BioX Interdisciplinary Initiatives Program, NHMRC Australia Fellowship, and Kwanjeong Educational Foundation (KEF). The authors thank GE Healthcare and Kris Thielemans for their loan and support of the Respiratory Gating Toolbox (RGT) software for 4D-PET image reconstruction. Siddharth Gopalan, Eric Johnston, Melody Chung, and Neville Eclov are thanked for their contributions to the development of the audiovisual biofeedback system and the patient data acquisition. Julie Baz improved the clarity of the manuscript. The authors are grateful to Dr. Rie Van Eyben, a statistician at Stanford, for the statistical analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2016.05.016.

References

[1] Nehmeh SA, Erdi YE. Respiratory motion in positron emission tomography/computed tomography: a review. Semin Nucl Med 2008;38:167–76.
[2] Yamamoto T, Langner U, Loo BW, Shen J, Keall PJ. Retrospective analysis of artifacts in four-dimensional CT images of 50 abdominal and thoracic radiotherapy patients. Int J Radiat Oncol Biol Phys 2008;72:1250–8.
[3] Yang J, Yamamoto T, Cho B, Seo Y, Keall PJ. The impact of audio-visual biofeedback on 4D PET images: results of a phantom study. Med Phys 2012;39:1046–57.
[4] Venkat RB, Sawant A, Suh Y, George R, Keall PJ. Development and preliminary evaluation of a prototype audiovisual biofeedback device incorporating a patient-specific guiding waveform. Phys Med Biol 2008;53:N197–208.
[5] Kim T, Pollock S, Lee D, O'Brien R, Keall P. Audiovisual biofeedback improves diaphragm motion reproducibility in MRI. Med Phys 2012;39:6921–8.
[6] Lu W, Neuner GA, George R. Audio-visual biofeedback does not improve the reliability of target delineation using maximum intensity projection in 4-dimensional computed tomography radiation therapy planning. Int J Radiat Oncol Biol Phys 2014;88:229–35.
[7] Nehmeh SA, Erdi YE, Ling CC, et al. Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer. Med Phys 2002;29:366–71.
[8] Rietzel E, Pan T, Chen GT. Four-dimensional computed tomography: image formation and clinical protocol. Med Phys 2005;32:874–89.
[9] Pan T, Lee TY, Rietzel E, Chen GT. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. Med Phys 2004;31:333–40.
[10] Nehmeh SA, Erdi YE, Ling CC, et al. Effect of respiratory gating on quantifying PET images of lung cancer. J Nucl Med 2002;43:876–81.
[11] Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med Phys 2006;33:3874–900.
[12] Greco C, Rosenzweig K, Cascini GL, Tamburrini O. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). Lung Cancer 2007;57:125–34.
[13] Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 2010;37:2165–87.
[14] Cui G, Jew B, Hong JC, Johnston EW, Loo BW, Maxim PG. An automated method for comparing motion artifacts in cine four-dimensional computed tomography images. J Appl Clin Med Phys 2012;13:3838.
[15] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 1989:319–40.
[16] Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989;35:982–1003.
[17] Brédart A, Morvan E, Savignoni A, Giraud P. Patient's perception of care quality during radiotherapy sessions using respiratory gating techniques: validation of a specific questionnaire. Cancer Invest 2011;29:145–52.
[18] George R, Chung TD, Vedam SS, et al. Audio-visual biofeedback for respiratory-gated radiotherapy: impact of audio instruction and audio-visual biofeedback on respiratory-gated radiotherapy. Int J Radiat Oncol Biol Phys 2006;65:924–33.
[19] Pollock S, Lee D, Kim T, et al. Respiratory guidance for lung cancer patients: an investigation of audiovisual biofeedback training and effectiveness. Med Phys 2013;40:183.
[20] Persson GF, Nygaard DE, Olsen M, et al. Can audio coached 4D CT emulate free breathing during the treatment course? Acta Oncol 2008;47:1397–405.
[21] Goossens S, Senny F, Lee JA, Janssens G, Geets X. Assessment of tumor motion reproducibility with audio-visual coaching through successive 4D CT sessions. J Appl Clin Med Phys 2014;15:4332.