From Tissue to Sound: Model-Based Sonification of Medical Imaging

Sasan Matinfar[1,2]([✉]), Mehrdad Salehi[1], Shervin Dehghani[1,2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP),
Technical University of Munich, 85748 Munich, Germany
`sasan.matinfar@tum.de`
[2] Nuklearmedizin rechts der Isar, Technical University of Munich,
81675 Munich, Germany

Abstract. We introduce a general design framework for the interactive sonification of multimodal medical imaging data. The proposed approach operates on a physical model that is generated based on the structure of anatomical tissues. The model generates unique acoustic profiles in response to external interactions, enabling the user to learn about how the tissue characteristics differ from rigid to soft, dense to sparse, structured to scattered. The acoustic profiles are attained by leveraging the topological structure of the model with minimal preprocessing, making this approach applicable to a diverse array of applications. Unlike conventional methods that directly transform low-dimensional data into global sound features, this approach utilizes unsupervised mapping of features between an anatomical data model and a sound model, allowing for the processing of high-dimensional data. We verified the feasibility of the proposed method with an abdominal CT volume. The results show that the method can generate perceptually discernible acoustic signals in accordance with the underlying anatomical structure. In addition to improving the directness and richness of interactive sonification models, the proposed framework provides enhanced possibilities for designing multisensory applications for multimodal imaging data.

Keywords: Sonification · Auditory Display · Auditive Augmented Reality · Model-based Sonification · Physical Modeling Sound Synthesis

1 Introduction

Carrying out a surgical procedure requires not only spatial coordination but also the ability to maintain temporal continuity during time-critical situations, underscoring the crucial importance of auditory perception due to the temporal

nature of sound. Perceptual studies have shown that stimuli in different sensory modalities can powerfully interact under certain circumstances, affecting perception or behavior [1]. In multisensory perception, sensory experiences are integrated when they occur simultaneously, i.e., when complementary or redundant signals are received from the same location at the same time. Previous studies have shown that combining independent but causally correlated sources of information facilitates information processing [2]. It has been demonstrated that multisensory integration, particularly auditory cues embedded in complex sensory scenes, improves performance on a wide range of tasks [3]. Biologically, humans have evolved to process spatial dimensions visually [4]. At the same time, their auditory system, which works in an omnidirectional manner, helps to maintain a steady pace in dynamic interaction, facilitating hand-eye coordination. It has been shown that bimanual coordination augmented with auditory feedback, compared to visual, leads to enhanced activation in brain regions involved in motor planning and execution [5].

However, the integration of auditory systems with visual modalities in biomedical research and applications has not been fully achieved. This could be due to the challenges involved in providing perceptually unequivocal and precise sound while reflecting high-resolution and high-dimensional information, as available in medical data. This paper proposes a novel approach, providing an intelligent modeling as a design methodology for interactive medical applications. Considering the limitations of the state-of-the-art, we investigate potential design approaches and the possibility of establishing a new research direction in sonifying high-resolution and multidimensional medical imaging data.

2 State-of-the-Art of Sonification in the Medical Domain

Sonification is the data-dependent generation of sound, if the transformation is systematic, objective and reproducible, so that it can be used as scientific method [6,7]. Researchers have investigated the impact of sound as an integral part of the user experience in interactive systems and interfaces [8]. In this form of interaction design, sound is used to create rich, dynamic, and immersive user experiences and to convey information, provide feedback, and shape user behavior. The use of sonification has expanded in recent years to a wide range of tasks, such as navigation, process monitoring, data exploration, among others [7]. A variety of sonification techniques for time-indexed data such as audification, parameter-mapping sonification (PMSon), and model-based sonification (MBS) translate meaningful data patterns of interaction into perceptual patterns, allowing listeners to gain insight into those patterns and be aware of their changes.

In PMSon, data features are input parameters of a mapping function to determine synthesis parameters. PMSon allows explicit definition of mapping functions in a flexible and adaptive manner, which has made it the most popular approach for medical applications. Medical sonification, as demonstrated by pioneering works such as [9–12], has predominantly used PMSon for sonifying the position or state of surgical instruments with respect to predetermined

structures or translating spatial characteristics of medical imaging data into acoustic features. Research in surgical sonification [13,14] has mainly focused on image-guided navigation, showing significant results in terms of accuracy and facilitating guidance. Fundamental research behind these studies has aimed to expand mapping dimensionality beyond 1D [15] and 2D [16,17] up to three orthogonal dimensions [18]. A sonification method combining two orthogonal mappings in two alignment phases has been proposed in [19] for the placement of pedicle screws with four degrees of freedom. In addition to expanding the data bandwidth, there have been studies [11,12,14,20,21] which address the issues of integrability and pleasantness of the resulting sound signal. This is arguably one of the most significant challenges in surgical sonification, which still requires further research.

Medical sonification has shown great potential, despite being a relatively new field of research. However, there are limitations in terms of design and integrability. The tedious and case-specific process of defining a mapping function that can be perceptually resolved, even in low dimensional space, poses a significant challenge, rendering the achievement of a generalized method nearly impossible. Consequently, practitioners endeavor to minimize data complexity and embed data space into low dimensional space, giving rise to sonification models that offer only an abstract and restricted understanding of the data, which is inadequate in the case of complex medical data. Although these methods achieve adequate perceptual resolution and accuracy, they still are not optimized for integration into the surgical workflow and demonstrate low learning rates, which also demand increased cognitive load and task duration. Furthermore, fine-tuning these models involves a considerable amount of artistic creativity in order to avoid undesirable effects like abrasiveness and fatigue.

3   Towards Model-Based Sonic Interaction with Multimodal Medical Imaging Data

The need for an advanced multisensory system becomes more critical in scenarios where the anatomy is accessed through limited means, as in minimally invasive surgery. In such cases, an enriched auditory feedback system that conveys precise information about the tissue or surrounding structures, capturing spatial characteristics of the data (such as density, solidity, softness, or sparsity), can enhance the surgeon's perception and, on occasion, compensate for the lack of haptic feedback in such procedures. Embedding such use cases into a low-dimensional space is not feasible. In the same way that tapping on wood or glass imparts information about the object's construction, an effective sonification design can convey information in a manner that is easily interpreted with minimal cognitive effort. Intuitive embodiment of information in auditory feedback, facilitates the learning process to the extent that subconscious association of auditory cues with events can be achieved. Despite the fact that anatomical structures do not produce sound in human acoustic ranges, and surgeons do not have a direct perceptual experience of them, these structures still adhere to

the physical principles of dynamics. The hypothesis of this paper is that sounds based on the principles of physics are more straightforward to learn and utilize due to their association with real-world rules.

MBS is a technique that utilizes mathematical models to represent data and then convert those models into audible sounds. This approach is often used in scientific or data analysis applications such as clustering [22] or data scanning [23], where researchers aim to represent complex data sets or phenomena through sound. MBS employs a mapping approach that associates data features with the features of a sound-capable system, facilitating generalization across different cases. MBS is often coupled with physical modeling synthesis that uses algorithms to simulate the physical properties of real-world instruments and generate sound in a way that mimics the behavior of physical objects [24–26]. This approach is often used to create realistic simulations of acoustic instruments or explore new sonic possibilities that go beyond the limitations of traditional instruments. Sound is a physical phenomenon that arises from the vibration of objects in a medium, such as air, resulting in a complex waveform composed of multiple frequencies at different amplitudes. This vibrational motion can be generated by mechanical impacts, such as striking, plucking, or blowing, applied to a resonant object capable of sustaining vibration. Physical modeling has been previously introduced to the biomedical research community as a potential solution for the issue of unintuitive sounds, as reported in [21]. However, as mentioned earlier, it is limited by the shortcomings of the current state-of-the-art, particularly with respect to reduced data dimensionality.

Physical modeling can be approached using the mass interaction method, which is characterized by its modular design and the capacity to incorporate direct gestural interaction, as demonstrated in [27–29]. This approach offers the advantage of describing highly intricate virtual objects as a construction of elementary physical components. Additionally, such an approach is highly amenable to the iterative and exploratory design of "physically plausible" virtual objects, which are grounded in the principles of Newtonian physics and three laws of motion but are not necessarily limited to the mechanical constraints of the physical world.

Contribution

This paper presents a novel approach for transforming spatial features of multimodal medical imaging data into a physical model that is capable of generating distinctive sound. The physical model captures complex features of the spatial domain of data, including geometric shapes, textures, and complex anatomical structures, and translates them to sound. This approach aims to enhance experts' ability to interact with medical imaging data and improve their mental mapping regarding complex anatomical structures with an unsupervised approach. The unsupervised nature of the proposed approach facilitates generalization, which enables the use of varied input data for the development of versatile sound-capable models.

4   Methods

The proposed method involves multimodal imaging data serving as input, a topology to capture spatial structures, and an interaction module to establish temporal progress. We consider medical imaging data as members of a $\mathbb{R}^{d+m}$ space, where d is the data dimensionality in space domain, and m the dimension of measured features by the imaging systems. A sequence of physics-based sound signals, S, expressed as

$$S = \sum_{i=0}^{n} P_i(act(pos_i, \vec{F}_i), f(A_i), T_i), \tag{1}$$

can be achieved by the sum of n excitations of a physical model P with user interaction act at the position pos with applied force of $\vec{F}$, where f transfers a region of interest (RoI) $A \in \mathbb{R}^{d+m}$ to parameters of the physical model P using the topology matrix T. These components are described in detail in Sect. 4.1. Figure 1 provides an illustration of the method overview and data pipeline.
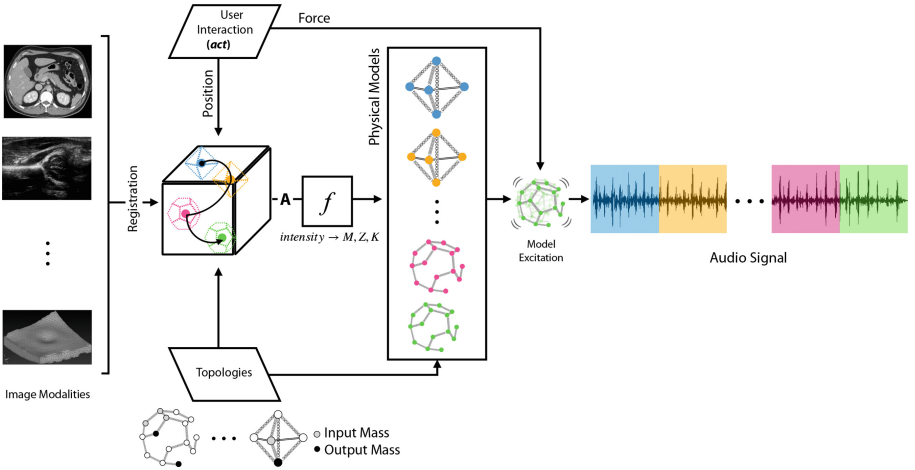

Fig. 1. The diagram shows the transformation of multimodal medical imaging data to sound. The registration method can be arbitrarily chosen depending on the application and data modalities.

4.1   Physical Model

The mass-interaction physics methodology [28] allows the formulation of physical systems, such as the linear harmonic oscillator, which comprise two fundamental constituents: masses, representing material points within a 3D space, with corresponding inertial behaviors, and connecting springs, signifying specific types of physical couplings such as viscoelastic and collision between two mass elements.

Implementation of a Mass-Interaction System in Discrete Time. To represent and compute discretized modular mass-interaction systems, a widely used method involves applying a second-order central difference scheme to Newton's second law, which states that force $\vec{F}$ is equal to mass m times acceleration a, or the second derivative of its position vector x with respect to time t.

The total force exerted by the dampened spring, denoted as $\vec{F}(t_n) = \vec{F}_s(t_n) + \vec{F}_d(t_n)$ at time t_n, where $\vec{F}_s$ represents the elastic force exerted by a linear spring (the interaction) with stiffness K, connecting two masses m_1, m_2 located at positions x_1, x_2, can be expressed using the discrete-time equivalent of Hooke's law. Similarly, the friction force $\vec{F}_d$ applied by a linear damper with damping parameter z can be derived using the Backward Euler difference scheme with the discrete-time inertial parameter $Z = z/\Delta T$. $\vec{F}(t_n)$ is applied symmetrically to each mass in accordance with Newton's third law: $\vec{F}_{2\rightarrow1}(t_n) = -\vec{F}(t_n)$ and $\vec{F}_{1\rightarrow2}(t_n) = +\vec{F}(t_n)$. The combination of forces applied to masses and the connecting spring yields a linear harmonic oscillator as described in

$$X(t_{n+1}) = (2 - \frac{K + Z}{M})X(t_n) + (\frac{Z}{M} - 1)X(t_{n-1}) + \frac{\vec{F}(t_n)}{M}, \qquad (2)$$

which is a fundamental type of the mass-interaction system. This system is achieved by connecting a dampened spring between a mass and a fixed point $x_1(t_n) = 0$.

A mass-interaction system can be extended to a physical model network with an arbitrary topology by connecting the masses via dampened springs. The connections are formalized as a routing matrix T of dimensions $r \times c$, where r denotes the number of mass points in the physical model network, and c represents the number of connecting springs, each having only two connections. A single mass can be connected to multiple springs in the network.

Interaction Module. An interaction module, *act*, excites the model P by applying force to one or more input masses of the model $I \in T^{r \times c}$. f maps the intensities of the input data to M, K, Z, and $\vec{F}$. Therefore, the input force is propagated through the network according to the Eq. 2 and observed by the output masses $O \in T^{r \times c}$.

To summarize, the output masses are affected by the oscillation of all masses activated in the model with various frequencies and corresponding amplitudes, resulting in a specific sound profile, i.e., tone color. This tone color represents the spatial structure and physical properties of the RoI, which is transformed into the features of the output sound. The wave propagation is significantly influenced by the model's topology and the structure of the inter-mass connections, which have great impact on activating spatial relevant features and sound quality.

4.2   Experiment and Results

The objective is to evaluate the feasibility of the proposed method in creating a model that is capable of generating discernible sound profiles in accordance

with the underlying anatomical structures. In particular, we aim to differentiate between the sound of a set of tissue types. Through empirical experimentation on an abdominal CT volume [31], we determined a model configuration that achieves stability and reduces noise to the desired level. The shape of the topology is a 3D cube of size 7 mm^3. The inter-mass connections are established at the grid spacing distance of 1 mm between each mass and its adjacent neighbor masses. All the masses located on the surface of the model are set as fixed points. To excite the model, equal forces are applied to the center of the model in a 3D direction and observed at the same position. The RoI is obtained by selecting 3D cubes with the same topology and size in the CT volume. The intensities in the RoI are transformed to define the model parameters. The spring parameters K and Z are derived by averaging the intensities of their adjacent CT voxels, by a linear mapping and M and $\vec{F}$ are set to constant values. We defined a sequence of RoI starting in the heart, passing through lung, liver, bone, muscle, and ending in the air in a 16-steps trajectory, as shown in Fig. 2. For visualizing the trajectory and processing image intensities, we used ImFusion Suite [32]. For generating physical models and generating sound, we used mass-interaction physical modelling library [29] for the Processing sketching environment: https://processing.org/. Visual demonstrations of the models, along with the corresponding sound samples, are provided in the supplementary material, along with additional explanations.

A mel spectrogram is a visual representation of the frequency content of an audio signal, where the frequencies are mapped to a mel scale, which is a perceptual frequency scale based on how humans hear sounds. Therefore, we used this representation to show the frequency content of the resulting sound of the trajectory, presented in Fig. 2.
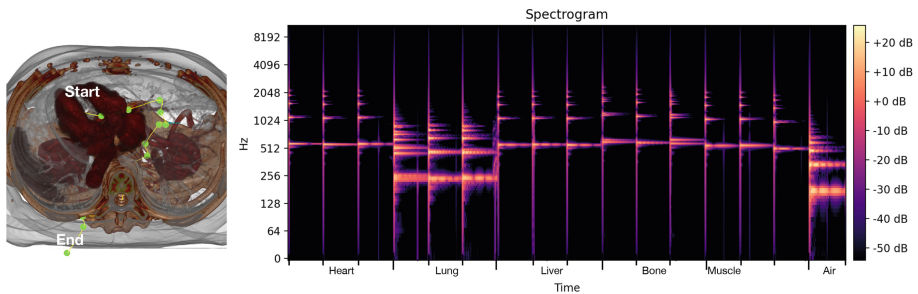

Fig. 2. The spectrogram illustrates the sound profiles of the tissues corresponding to the sequence of RoIs marked as yellow on the abdominal CT. The green spheres highlight the excitation points. (Color figure online)

5 Discussion and Conclusion

This paper introduced a general framework for MBS of multimodal medical imaging data. The preliminary study demonstrates perceptually distinguishable sound profiles between various anatomical tissue types. These profiles are achieved through a minimal preprocessing based on the model topology and a basic mapping definition. This indicates that the model is effective in translating intricate imaging data into a discernible auditory representation, which can conveniently be generalized to a wide range of applications. In contrast to the traditional methods that directly convert low-dimensional data into global sound features, this approach maps features of a data model to the features of a sound model in an unsupervised manner and enables processing of high-dimensional data.

The proposed method presents opportunities for several enhancements and future directions. In the case of CT imaging, it may be feasible to establish modality-specific generalized configurations and tissue-type-specific transfer functions to standardize the approach to medical imaging sonification. Such efforts could lead to the development of an auditory equivalent of 3D visualization for medical imaging, thereby providing medical professionals with a more immersive and intuitive experience. Another opportunity could be in the sonification of intraoperative medical imaging data, such as ultrasound or fluoroscopy, to augment the visual information that a physician would receive by displaying the tissue type or structure which their surgical instruments are approaching. To accommodate various application cases, alternative interaction modes can be designed that simulate different approaches to the anatomy with different tool materials. Additionally, supervised features can be integrated to improve both local and global perception of the model, such as amplifying regions with anomalies. Different topologies and configurations can be explored for magnifying specific structures in data, such as pathologies, bone fractures, and retina deformation. To achieve a realistic configuration of the model regarding the physical behavior of the underlying anatomy, one can use several modalities which correlate with physical parameters of the model. For instance, the masses can be derived by intensities of CT and spring stiffness from magnetic resonance elastography, both registered as a 3D volume.

An evaluation of the model's potential in an interactive setting can be conducted to determine its impact on cognitive load and interaction intuitiveness. Such an assessment can shed light on the practicality of the model's application by considering the user experience, which is influenced by cognitive psychological factors. As with any emerging field, there are constraints associated with the methodology presented in this paper. One of the constraints is the requirement to manually configure the model parameters. Nevertheless, a potential future direction would be to incorporate machine learning techniques and dynamic modeling to automatically determine these parameters from underlying physical structure. For instance, systems such as [30] can provide a reliable reference for such investigations. Considering the surgical auditory scene understanding is vital when incorporating the sonification model into the surgical workflow. Future studies

can investigate this by accounting for realistic surgical environments, including existing sound sources, and considering auditory masking effects.

References

1. Ernst, M.O., Di Luca, M.: Multisensory perception: from integration to remapping. In: Sensory Cue Integration, pp. 224–250 (2011)
2. Shams, L., Seitz, A.R.: Benefits of multisensory learning. Trends Cogn. Sci. **12**(11), 411–417 (2008)
3. Van der Burg, E., Olivers, C.N., Bronkhorst, A.W., Theeuwes, J.: Audiovisual events capture attention: evidence from temporal order judgments. J. Vis. **8**(5), 2 (2008)
4. Middlebrooks, J.C., Green, D.M.: Sound localization by human listeners. Annu. Rev. Psychol. **42**(1), 135–159 (1991)
5. Ronsse, R., et al.: Motor learning with augmented feedback: modality-dependent behavioral and neural consequences. Cereb. Cortex **21**(6), 1283–1294 (2011)
6. Hermann, T.: Taxonomy and definitions for sonification and auditory display. In: International Community for Auditory Display (2008)
7. Hermann, T., Hunt, A., Neuhoff, J.G.: The Sonification Handbook, vol. 1. Logos Verlag, Berlin (2011)
8. Franinovic, K., Serafin, S. (eds.): Sonic Interaction Design. MIT Press, Cambridge (2013)
9. Wegner, C.M., Karron, D.B.: Surgical navigation using audio feedback. In: Medicine Meets Virtual Reality, pp. 450–458. IOS Press (1997)
10. Ahmad, A., Adie, S.G., Wang, M., Boppart, S.A.: Sonification of optical coherence tomography data and images. Opt. Express **18**(10), 9934–9944 (2010)
11. Hansen, C., et al.: Auditory support for resection guidance in navigated liver surgery. Int. J. Med. Robot. Comput. Assist. Surg. **9**(1), 36–43 (2013)
12. Matinfar, S., et al.: Surgical soundtracks: towards automatic musical augmentation of surgical procedures. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 673–681. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_76
13. Black, D., Hansen, C., Nabavi, A., Kikinis, R., Hahn, H.: A survey of auditory display in image-guided interventions. Int. J. Comput. Assist. Radiol. Surg. **12**, 1665–1676 (2017). https://doi.org/10.1007/s11548-017-1547-z
14. Joeres, F., Black, D., Razavizadeh, S., Hansen, C.: Audiovisual AR concepts for laparoscopic subsurface structure navigation. In: Graphics Interface 2021 (2021)
15. Parseihian, G., Gondre, C., Aramaki, M., Ystad, S., Kronland-Martinet, R.: Comparison and evaluation of sonification strategies for guidance tasks. IEEE Trans. Multimedia **18**(4), 674–686 (2016)
16. Ziemer, T., Black, D., Schultheis, H.: Psychoacoustic sonification design for navigation in surgical interventions. In: Proceedings of Meetings on Acoustics, vol. 30, no. 1, p. 050005. Acoustical Society of America (2017)
17. Ziemer, T., Schultheis, H., Black, D., Kikinis, R.: Psychoacoustical interactive sonification for short range navigation. Acta Acust. Acust. **104**(6), 1075–1093 (2018)
18. Ziemer, T., Schultheis, H.: Psychoacoustical signal processing for three-dimensional sonification. Georgia Institute of Technology (2019)
19. Matinfar, S., et al.: Sonification as a reliable alternative to conventional visual surgical navigation. Sci. Rep. **13**(1), 5930 (2023). https://www.nature.com/articles/s41598-023-32778-z

20. Matinfar, S., Hermann, T., Seibold, M., Fürnstahl, P., Farshad, M., Navab, N.: Sonification for process monitoring in highly sensitive surgical tasks. In: Proceedings of the Nordic Sound and Music Computing Conference 2019 (Nordic SMC 2019) (2019)

21. Roodaki, H., Navab, N., Eslami, A., Stapleton, C., Navab, N.: SonifEye: sonification of visual information using physical modeling sound synthesis. IEEE Trans. Vis. Comput. Graph. **23**(11), 2366–2371 (2017)

22. Hermann, T., Ritter, H.: Listen to your data: model-based sonification for data analysis. In: Advances in Intelligent Computing and Multimedia Systems, vol. 8, pp. 189–194 (1999)

23. Bovermann, T., Hermann, T., Ritter, H.: Tangible data scanning sonification model. Georgia Institute of Technology (2006)

24. Smith, J.O.: Physical modeling using digital waveguides. Comput. Music. J. **16**(4), 74–91 (1992)

25. Cook, P.R.: Physically informed sonic modeling (PhISM): synthesis of percussive sounds. Comput. Music. J. **21**(3), 38–49 (1997)

26. Smith, J.O.: Physical audio signal processing: for virtual musical instruments and audio effects. W3K Publishing (2010)

27. Leonard, J., Cadoz, C.: Physical modelling concepts for a collection of multisensory virtual musical instruments. In: New Interfaces for Musical Expression 2015, pp. 150–155 (2015)

28. Villeneuve, J., Leonard, J.: Mass-interaction physical models for sound and multisensory creation: starting anew. In: Proceedings of the 16th Sound & Music Computing Conference, pp. 187–194 (2019)

29. Mass Interaction Physics in Java/Processing Homepage. https://github.com/mi-creative/miPhysics_Processing. Accessed 4 Mar 2023

30. Illanes, A., et al.: Novel clinical device tracking and tissue event characterization using proximally placed audio signal acquisition and processing. Sci. Rep. **8**(1), 12070 (2018)

31. Luo, X., et al.: WORD: a large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from CT image. Med. Image Anal. **82**, 102642 (2022)

32. Zettinig, O., Salehi, M., Prevost, R., Wein, W.: Recent advances in point-of-care ultrasound using the *ImFusion Suite* for real-time image analysis. In: Stoyanov, D., et al. (eds.) POCUS/BIVPCS/CuRIOUS/CPM -2018. LNCS, vol. 11042, pp. 47–55. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01045-4_6