International Research Journal of Computer Science (IRJCS)
Issue 4, Volume 2 (April 2015)
ISSN: 2393-9842
www.irjcs.com

PetaFlow: a global computing-networking-visualisation unit with social impact

A. Ancel[1], I. Assenmacher[1], K. Baba[7], J. Cisonni[2], Y. Fujiso[3], P. Goncalves[1,4], M. Imbert[4], K. Koyamada[5], P. Neyron[1], K. Nozaki[2], H. Ohsaki[2] A.C. Orgerie[4,8], X. Pelorson[3], B. Raffin[1,3], N. Sakamoto[5], E. Sakane[6], S. Wada[2], S. Shimojo[2], A. Van Hirtum[3,8*]

[1]*INRIA - France,* [2]*Osaka University - Japan,* [3]*Grenoble University Alpes- France,*
[4]*ENS Lyon - France,* [5]*Kyoto University - Japan,* [6]*NII, Japan -* [7]*Kogakuinoto University - Japan,* [8]*CNRS, France*

Abstract— The PetaFlow application aims to contribute to the use of high performance computational resources for the benefit of society. To this goal the emergence of adequate information and communication technologies with respect to high performance computing-networking-visualisation and their mutual awareness is required. The developed technology and algorithms are presented and applied to a real global peta-scale data intensive scientific problem with social and medical importance, i.e. human upper airflow modelling.

Keywords— large distance network, scientific visualisation, peta-scale data, high performance computation, remote health care

I. INTRODUCTION

It is no falsehood to state that `current society and science attempts to deal with increasing amounts of data'. Today, peta-scale data are commonly gathered as well as generated thanks to the continuous development of measurement technologies and computational resources in diverse fields of science and society. Efficient processing or generation of peta-scale data requires high performance computational (HPC) resources which should be made remotely accessible through long-distance high performance networking and might be represented thanks to interactive scientific visualisation. Consequently, generation or processing of peta-scale data benefits from the emergence of adequate `Information and communication technologies (ICT)' with respect to high performance `computing-networking-visualisation' and their mutual `awareness'. In the PetaFlow-application, it is aimed to develop and validate such ICT solutions using a transnational high-speed research network between Japan and France connecting `GRID5000' (France) to the `Naregi' (Japan) testbed. Data-transfer protocols are aimed to be validated on data obtained for a real scientific problem involving peta-scale data. Due to the medical relevance as well as basic scientific interest, peta-scale data are obtained from HPC Computational Fluid Dynamics (CFD) simulations on a vector supercomputer (NEC SCX9 Japan) aiming to predict airflow through the upper airways. In addition, CFD simulation outcome is used as an input for aero-acoustic computations (CAA) for prediction of noise production. The outcome of CFD and CAA simulations is validated on flow and noise measurements on a suitable experimental setup. Besides the international transfer of the generated peta-scale data, scientific visualisation of peta-scale data is aimed on a single PC as well as on a tiled display wall for 3D interactive reconstruction of the flow and noise data. In summary, PetaFlow aims to contribute to the state-of-the-art of HPC, networking, scientific visualisation and their mutual interactions for peta-scale data, while at the same time it is aimed to contribute to basic research in the fields of CFD and CAA applied to flow through the upper airways as illustrated in Fig. 1. The results presented in the following sections are due to joining efforts and resources of the French-Japanese consortium gathering specialists in networking, middleware, scientific visualisation, HPC and upper airway flow modelling and noise production.

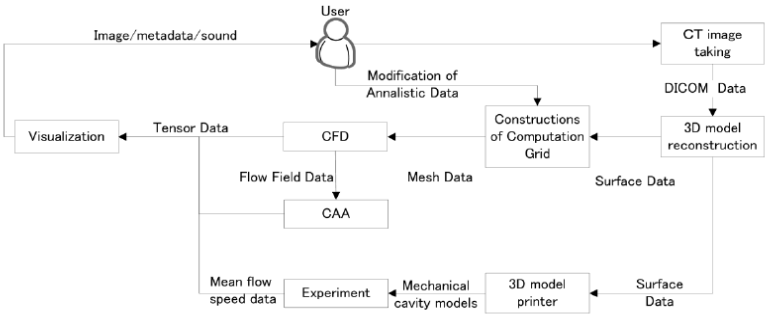


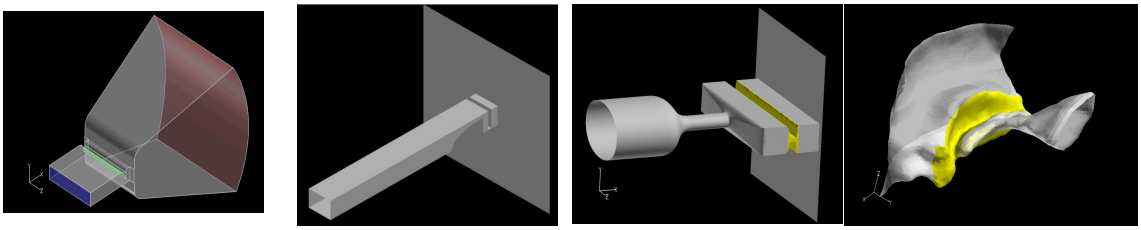
Fig. 1 Overview of the ICT solutions within the PetaFlow application: flow through the oral tract for dental practice

II. PETA-SCALE DATA GENERATION

A. *Real-life Application: Human Upper Airway Flow*

The flow through the human upper airways is subject to basic research due to the complexity and continuous variation of flow and boundary conditions: time-varying geometry, time-varying flow properties and changing tissue-properties (e.g. soft at the tongue and rigid at the teeth). Besides the scientific interest, upper airflow modelling is of social importance since it concerns every one of us. Airflow through the upper airways is crucial to life essential actions, such as breathing or coughing, or common social actions, such as speech production. In this framework, accurate prediction of upper airway flow is in particular important when dealing with upper airway pathologies. A few common examples are obstructive sleep apnea (5% of adult population), vocal folds polyps (common among e.g. teachers due to intensive professional voice use), cancer of the tongue/larynx requiring surgical interventions or common interventions at the teeth. Consequently, accurate upper airway flow modelling contributes to the development of tools aiming to improve basic health care and daily life comfort with respect to e.g. surgical planning, speech therapy, prosthesis design (teeth or larynx) or detection of the breathing zone for applications dealing with indoor air quality for follow-up or prevention of chronic diseases such as asthma. It is obvious that an accurate prediction requires appropriate computational resources. High performance computing (HPC) is needed in particular to predict fricative noise due to the requested accuracy of the flow field (up to 16kHz) and to obtain an accurate prediction within a `reasonable' amount of time. The reported simulation time for a grid size of $3x10^6$ number of cells on a 16-node Opteron cluster running Linux/MPI was reported 14 days [7]. Clearly, such computational times prevent systematic evaluation of configuration parameters or model approaches and parameters. Implementation of HPC in daily clinical practice or scientific environment requires access to remote computational resources and `on-line' visualisation of results in an appropriate interactive way. Consequently, the ubiquitous use and implementation of high performance computations in society and science request the development of adequate TCI in order to manage long distance data transfer and interactive visualisation. It is aimed to perform advanced grid computation of respiratory airflow through the upper airways - from the glottal constriction, up to the teeth and beyond including the breathing zone and their merging - in order to contribute to the understanding of upper airway flow, to the development of accurate predictions for surgical interventions and prosthesis design and to develop the necessary ICT solutions enabling long-distance data transfer in combination with interactive scientific visualisation.

Upper airway flow is the result of complex and varying geometry and flow conditions. Geometries with `in-vivo' complexity during sibilant utterances can be obtained from the oral cavity shape and the incisors estimated by volume rendering of Cone Beam CT scans (512 slices by 512x512 pixels within 18 seconds) [6]. Simplified geometries are defined based on `in-vivo' observations (Fig. 2) in order to allow the impact of a particular geometrical parameter and to facilitate experimental `in-vitro' validation of the simulated flow field. Since the presence of a teeth as an obstacle in the flow is assumed to be a strong acoustic source, simplified geometries focus on the teeth shaped obstacle in several configurations. Validation of the simulation results is a crucial step before application to the field of medical health care. Due to the required high performance computational resources systematic studies of flow and geometry parameters - in order to understand their influence on upper airway flow and related quantities, such as sound production - are despite their potential applications and scientific interest currently missing or few.



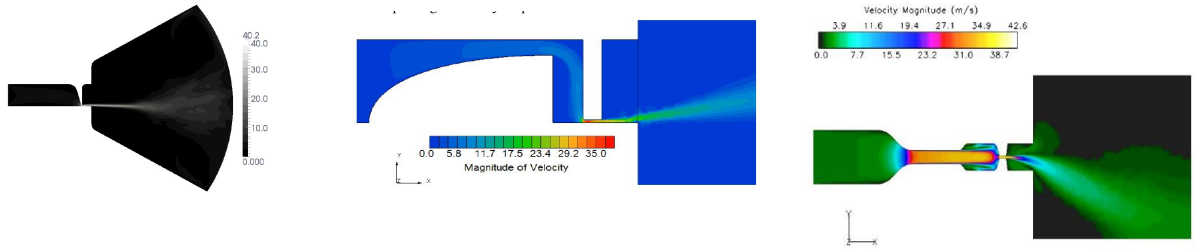
(a) Incisor and free field [2] (b) Tongue and incisor [3] (c) Tongue and incisor [5] (d) Human [6]

Fig. 2 Examples of oral tract geometries during sibilant /s/ utterance with increasing complexity studied in PetaFlow

B. *Flow and Acoustic Simulated Data*

A systematic study of airflow through simplified oral tract geometries shown in Fig. 2 as function of geometrical parameters is performed as illustrated in Fig. 3 [5,2,4,8,9]. The flow domain consists of an oral tract geometry extended with an outlet cavity (Fig.2a) representing free field of sufficient length to study jet development downstream the model. Spatial discretization of the whole flow domain is carried out resulting in an structured (Gridgen, Pointwise Inc.) mesh of hexahedral 8-node finite elements or an unstructured (SC-Tetra v8, Cradle Co.) mesh of tetrahedral finite volumes.

The mesh is refined in directions and regions a-priori associated with rapid flow variations such as the main streamwise direction, constricted portions, boundary layers and regions containing jet development. Resulting mesh sizes vary from $5x10^6$ up to $10x10^6$ elements corresponding to 355MB up to 980MB. Boundary conditions (3MB up to 9MB) are imposed corresponding to laminar uniform inflow at moderate Reynolds number (maximum 4000), rigid channel walls with no-slip condition and static zero pressure at the channel outlet in combination with no back flow.

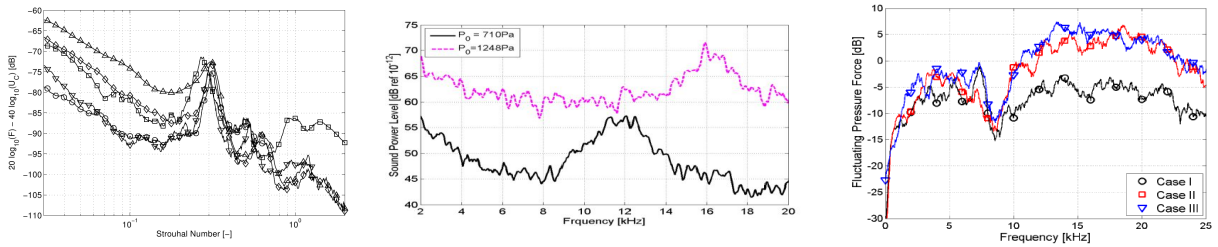


(a) Incisor and free field [2] (b) Tongue and incisor [4] (c) Tongue and incisor [5]

Fig. 3 Examples of average velocity [m/s] for simplified geometries shown in Fig. 2

The airflow is simulated with Large Eddy Simulation (LES) [10] of the governing Navier-Stokes equation for incompressible unsteady flows implemented in Front Flow Blue [11]. The spatially filtered momentum equation and continuity equation are solved over the discretised flow domain so that sought flow quantities are described as a combination of a resolvable large scale part and a subgrid-scale part. The subgrid scale part is modelled using dynamic Smagorinsky turbulence model for which the eddy viscosity is determined locally by Lilly's least square procedure based on the element volume [12]. To ensure numerical stability the value of the time step used in the numerical simulation depends on the mesh size in order to satisfy the Courant-Friedrichs-Lewy (CFL) condition. At first, the inlet velocity is gradually increased during 10^4 time steps. Then the flow field is simulated with a constant inlet velocity during 100ms up to 600ms depending on the constriction aperture at the incisors. Flow simulations are performed using NEC SX9 supercomputer, allowing the computation of 10000 times steps in approximately 13 hours. The total number of time steps varied between $2x10^4$ (large constriction) and $3x10^5$ (small constriction). The flow field raw data yield up to 500MB per time step depending on the geometry. Consequently, systematic study of an obstacle inserted in different channel configurations as shown in Fig. 2 as well as for different inflow conditions result in a huge amount, i.e. peta-scale, of localised data.

For moderate Reynolds number (maximum 4000) interactions of the flow with the surrounding walls, i.e. fluctuations of the pressure forces on the walls of the teeth-shaped obstacle, are assumed to be the main source of flow-induced noise due to the associated value of the Mach number (<0.2). Thus, the flow-induced noise as well as its spectral properties can be estimated from the surface pressure predicted by the flow simulation resulting in dipole sources on the solid boundary [13]. Computed noise spectra are illustrated in Fig. 4.



(a) Incisor and free field [2] (b) Tongue and incisor [4] (c) Tongue and incisor [5]

Fig. 4 Examples of fluctuating pressure force exerted on the walls surface for simplified geometries shown in Fig. 2 as a function of: a) aperture degree at the obstacle 30% (triangle), 20% (diamond), 10% (square) and 6% (circle), b) inlet pressure and c) centreline symmetry: no offset (I), constriction offset (II) and cavity offset (III)

Validation of flow and acoustic quantities is performed by mounting mechanical replicas of the geometries shown in Fig. 2 in a suitable experimental setup so that inflow conditions can be controlled and physical quantities can be measured. Experimental validation and characterisation is assessed by applying hot film anemometry [5], spatial pressure sampling [3], acoustic measurements [14,15] and flow visualisation techniques [8]. The same way as for data obtained from numerical simulation a systematic validation of geometrical configurations and inflow conditions results in a large amount of localised data, which increases considerably in case the flow field is sampled using visualisation techniques.

C. CFD/CAA: a well designed HPC approach

A vector supercomputer is used to perform the simulations reported on in the pervious section. Such a powerful computational environment is not general available and can be hard to adapt to in case coupled simulations need to be done such as a for instance a coupled simulation involving computational fluid dynamics (CFD) and computational aero-acoustics (CAA). In order to optimise the performance in terms of total simulation execution time, a scheduling system has been proposed [16] for a distributed computational environment based on network monitoring data accounting for computational fluid dynamics (CFD), computational aero-acoustics (CAA), data transfer among computer nodes and user interaction needed to visualise intermediate results.

III. EXTRA LONG-DISTANCE PETA-SCALE DATA TRANSFER

Using the www and scp protocols to transfer simulated data between Japan and France results in a data transfer rate of 150 up to 300 kB/s. So that about 20 minutes are required to transfer a data file containing one computed time step and 50 minutes to transfer a grid file. For 100 computed time steps (31GB) this results in a transfer time of 37 hours and 30 minutes. Consequently, in case a remote supercomputer is used for a real application resulting in peta-scale data, as is the case in PetaFlow, data transfer limits the computational efficiency and the interaction of the user with the data.

A. An international network testbed

The PetaFlow network testbed is a layer-2 virtual private network(VPN). It has been developed from the NAREGI-Grid5000 network testbed (2006-2009) and constructed through a collaboration among SINET, JGN-X, RENATER, GEANT, and MAN LAN. Fig. 5 shows the topology of the PetaFlow network testbed. On the Japanese side, the network is composed of SINET and JGN-X networks, which are connected at Tokyo. The NII and Kyoto University connect with SINET, and Osaka University connects with JGN-X. The international network operated by SINET is used to connect the Japanese research foothold with Grid5000, and this network extends to MAN LAN (New York, USA) via GEANT (Europe). The Grid5000 backbone network (https://www.grid5000.fr) is provided by RENATER.

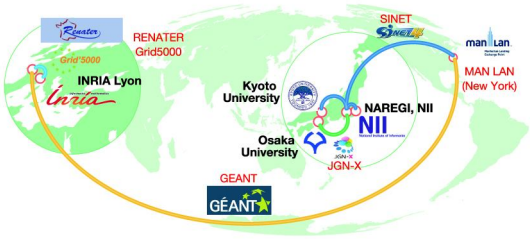


Fig. 5 Network topology in PetaFlow testbed

B. Networking, protocols and data transfer

In order to monitor the data transfer operated over the transnational link deployed between Japan and France and described in Section 3.1, we developed a synchronised capture system that was presented in [17]. We retained a hybrid solution that gathers the best of the two standard approaches: accuracy and high performance from hardware-based solutions by using programmable network equipment which will be responsible for the packet capture; and deployability and low cost from software-based solutions that will be used to develop drivers for these network cards. Actually, any NIC can be used to perform network metrology as demonstrated with the software-based solutions. Yet, non- specialized NICs require software implementations to handle packet capture and they offer a lower time-stamping accuracy. Programmable devices can present an interesting compromise between accuracy and scalability. The hybrid solutions presented in [17] rely on such programmable devices and come with firmware performing packet capture. However, these solutions which are not based on GPS synchronisation still lack accuracy to deal with high speed traffic. To circumvent this difficulty, we resorted to the Robust Absolute and Difference Clock (RADclock) [18]. RADclock is a system for network timing providing a difference clock to measure accurately the time elapsed between two events, and an absolute clock to get timestamps (like in the current system clocks). The timestamping process of RADclock is performed in the kernel. Unlike NTPd, the correction parameters are not directly applied to the system clock. They are instead used only when a timestamp is required, thus limiting the impact of frequent adjustments and inconsistencies. everal studies have shown the great robustness and accuracy of RADclock compared to other solutions like NTP or PTP [18,19,20,21]. Yet, it relies ultimately on NTP Stratum-1 servers, which are the reference clocks for this method. Fig. 6 presents the results on a capture lasting over 15 hours. It shows the estimation error for packets going from a server to a client. These packets are received at the same time by a DAG card and our NIC card based solution. RAD-DAG represents the data for our solution against the DAG data while SYS-DAG represents the data for the solution based on NTP against the DAG data. The real timestamp value is represented here by the DAG card. The ideal value is thus close to 0. The RADclock based solution is closer to the real value, which means it is more accurate. Moreover, as we can see from the distribution function, this solution is also less variable (showed by a factor of 5 on the interquartile range), thus more robust to changes in network traffic conditions than the NTP-based solution.

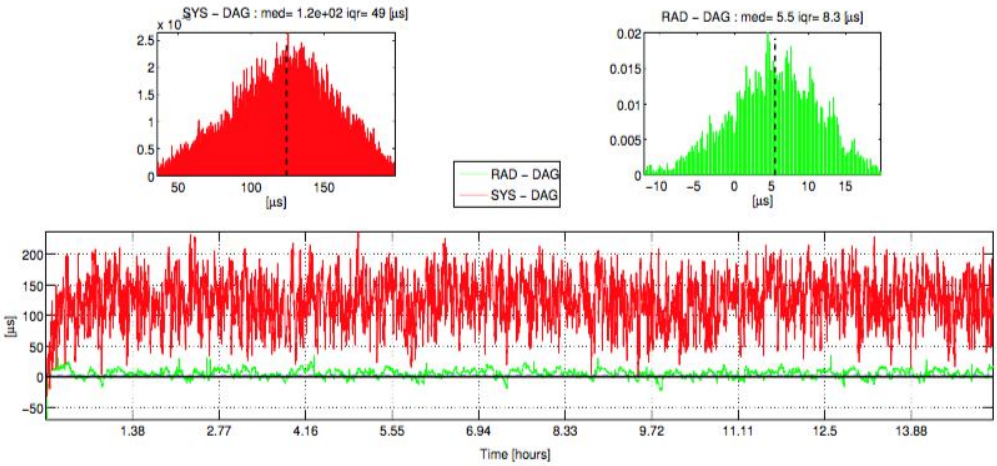


Fig. 6 Estimation error for packets going from server to client

IV. SCIENTIFIC VISUALISATION

The visualisation side is the last step in the PetaFlow application. Given the network constraints and the data scale resulting from the simulation, itis a challenging step.

A. *FlowVR*

To implement our work, we rely on the FlowVR data flow middleware, as presented by [22]. FlowVR enables to encapsulate existing codes in components, to interconnect them through data flow channels and to deploy them on distributed computing resources. FlowVR relieves the application developer from all networking and deployment details. The base entity, a component, is an autonomous process, potentially multi-threaded, that processes data coming from input ports and writes data on output ports. A component has no global insight on where the data comes from or goes to. The component programming interface is designed to limit code refactoring, and to ease the transformation of an existing code into a FlowVR component. The three main functions are:

- wait(): a blocking function call that waits for the availability of new messages on input ports,
- get(): retrieve a handle to access the message received at the previous wait() call on a given input port,
- put(): notify FlowVR that a new message on a given output port is ready for dispatch.

FlowVR manages data transfers. Intra-node communications between two components take place through a shared memory segment, avoiding copies. Once the sender has prepared the data in a shared memory segment, it simply handles a pointer to the destination that can directly access them. Inter-node communications extend this mechanism, FlowVR taking care of packing and transferring the data from the source shared memory segment to the destination shared memory segment. From the user point of view defining an application only consists in mapping components to nodes and connecting ports through a Python script. FlowVR is well suited for deploying parallel applications from cluster to transcontinental scale.

A. On-line visualisation

We proposed in [23] to use FlowVR combined with Particle-based Volume Rendering (PBVR) [24]. PBVR consists in generating sets of emissive opaque particles from a given volume dataset. It works in several steps. First, different sets of particles are generated in a cell-by-cell manner using a given transfer function and different random seeds. This process is typically done on the CPU side. The particle projection is then done on the GPU by projecting the different sets on the image plane and averaging the results. The number of particle sets used to generate an image is called the repeat level: lowering the repeat level increases the rendering speed, whereas increasing the repeat leve increases the rendering quality. We use FlowVR in [23] to propose two ways of visualising the result of PBVR rendering: Server and client side approaches. First, we consider a server entity that hosts the datasets and a client entity on which we want to visualise the data. The server side visualisation proposes to use the GPUs present on the server. The PBVR rendering is thus done on the server and the generated images are streamed to the visualisation client. The client side approach consists in generating the particles or reading them from files storing pre-computed particles on the server, and then stream them to the client so the final visualisation step can be done and the image displayed to the client.

V. CONCLUSIONS

A real-scale international long-distance application is presented involving large remote HPC datasets, computational resources and users. It is shown that the bottleneck for efficient usage of large-scale computational resources including interaction by scientific visualisation consists in network constraints. An important aspect of network constraint is due to its rigid management inherent to todays network strategies. Consequently, a major future challenge consists in the development and proof-of-concept of probabilistic strategies accounting for requirements of data, user interaction and physical network limitations which should enable a more flexible network usage with respect to data access. A second major challenge in case of HPC simulations relies in the further development of in-situ data analysis. Thirdly, the current application deals with upper airway flow and in particular the reconstruction of sound production for medical purposes such as the prediction of surgical interventions in case of cancer or denture prosthesis design. Obviously, a widespread usage of oral e-science benefits to society and results in large databases which can in turn be used for machine learning avoiding the tedious HPC computational approach. On the other hand in case diagnostic purposes are aimed, distributed sensor networks, to capture environmental air quality or/and acoustic data measured on patients, needs to be considered which in turn poses other requirements to the network strategy. Therefore, the feasibility of network protocols capable to deal with the combined requirements of 1) generation/interaction of large amounts of HPC data and 2) small amounts of data gathered from large distributed sensor networks is an interesting question for further research with many applications. In terms of an e-health application such a solution unites diagnostics and treatment, which is a major challenge for future ICT-health-technologies.

ACKNOWLEDGMENT

Authors thank ANR-JST project (France-Japan project PETAFLOW - ANR-09-BLAN-0376-02) for financial support.

REFERENCES

[1] Baba K, Cisonni J, Ebara Y, Gon ̧calves P, Grandchamp X, Kawamura T, Koyamada K, Nozaki K, Ohsaki H, Pelorson X, Primet P, Raffin B, Sakane E, Sakamoto N, Shimojo S, Van Hirtum A, Wada S. Petaflow: a project towards information and communication technologies in society. In: Proc. 1st workshop on High Speed Network and Computing Environ- ments for Scientific Applications (HSNCE 2010). Seoul, South-Korea; 2010:1–4.

[2] Cisonni J, Nozaki K, Van Hirtum A, Wada S. A parameterized geometric model of the oral tract for aeroacoustic simulation of the fricatives. Int J of Information and Electronics Engineering (IJIEE) 2011;1:223–228.

[3] Van Hirtum A, Pelorson X, Estienne O, Bailliet H. Experimental validation of flow models for a rigid vocal tract replica. J Acoust Soc Am 2011;130:2128.
[4] Cisonni J, Nozaki K, Van Hirtum A, Wada S. Physical analysis of the unvoiced fricatives production bases on a 2-articulators model. In: Proc. Japanese Fluid Mechanics Conf. Tokyo, Japan; 2011:1–4.
[5] Van Hirtum A, Grandchamp X, Pelorson X, Nozaki K, Shimojo S. LES and 'in vitro' experimental validation of flow around a teeth-shaped obstacle. Int J Applied Mech 2010;2(2):265–279.
[6] Nozaki K. Numerical simulation of sibilant [s] using the real geometry of a human vocal tract. Springer Berlin Heidelberg; 2010:.
[7] Ramsay G. The influence of constriction geometry on sound generation in fricative consonants. J Acoust Soc Am 2008;123:3579.
[8] Van Hirtum A, Grandchamp X, Cisonni J, Nozaki K, Bailliet H. Numer- ical and experimental exploration of flow through a teeth-shaped nozzle. Adv and Appl in Fluid Mech 2012;11:87–117.
[9] Cisonni J, Nozaki K, Van Hirtum A, Grandchamp X, Wada S. Numerical simulation of the influence of the orifice aperture on the flow around a teeth-shaped obstacle. Fluid Dyn Res 2013;45:1–19.
[10] Sagaut P, Germano M. Large Eddy Simulation for Incompressible Flows. New York: Spring Verslag; 2005.
[11] Guo Y, Kato C, Yamade Y. Basic features of the fluid dynamics simulation software FrontFlow/Blue. Japanese Society Fluids Mechanics 2006;58:11–15.
[12] Lilly D. A proposed modification of the Germano subgrid-scale closure method. Physics of Fluids 1992;4:633–635.
[13] Curle S. The influence of solid boundaries upon aerodynamic sound. Phil Trans R Soc Lond A 1955;231:505–514.
[14] Fujiso Y, Van Hirtum A, Nozaki K, Wada S. Experimental and numerical characterisation of aerodynamic noise applied to moderate Reynolds number airflow. In: Proc. 21st Int. Conf. on Acoustics. Montreal, Canada; 2013:1–9.
[15] Fujiso Y, Van Hirtum A, Nozaki K, Wada S. Study of unvoiced fricative speech production: influence of initial conditions on flow development. In: Proc. 21st Int. Conf. on Acoustics (ICA). Montreal, Canada; 2013:1– 9.
[16] Nozaki K, Ohsaki H, Tanaka M, Van Hirtum A, Sakane E. Stategic optimization of computation nodes allocation in a coupled simulation. In: Proc. 3th workshop on High Speed Network and Computing Environments for Scientific Applications (HSNCE 2012). Izmir, Turkey; 2012:.
[17] Orgerie AC, Gonc¸alves P, Imbert M, Ridoux J, Veitch D. Survey of net- work metrology platforms. In: 12th IEEE/IPSJ International Symposium on 12th IEEE/IPSJ SAINT, 3rd Workshop on High Speed Network and Computing Environments. Izmir (Turkey); 2012:.
[18] Veitch D, Ridoux J, Korada SB. Robust Synchronization of Absolute and Difference Clocks over Networks. IEEE/ACM Transactions on Networking 2009;17:417–430.
[19] Ridoux J, Veitch D. The Cost of Variability. In: Int. IEEE Symp. Precision Clock Synchronization for Measurement, Control and Communication (ISPCS'08). 2008:29–32.
[20] Ridoux J, Veitch D. Ten Microseconds Over LAN, for Free (Extended). IEEE Trans Instrumentation and Measurement (TIM) 2009;58(6):1841– 1848.
[21] Ridoux J, Veitch D, Broomhead T. The Case for Feed-Forward Clock Synchronization. IEEE/ACM Transactions on Networking 2012;20:231– 242.
[22] Allard J, Lesage JD, Raffin B. Modularity for Large Virtual Reality Applications. Presence: Teleoperators and Virtual Environments 2010;19:142–161.
[23] Lorant A, Ancel A, Zhao K, Sakamoto N, Koyamada K, Raffin B. Particle-based volume rendering of remote volume datasets using flowvr. In: AsiaSim. 2012:285–296.
[24] Sakamoto N, Nonaka J, Koyamada K, Tanaka S. Particle-based volume rendering. In: Visualization, 2007. APVIS '07. 2007 6th International Asia-Pacific Symposium on. 2007:129–132.