BlowClick: A Non-Verbal Vocal Input Metaphor for Clicking

Daniel Zielasko* Sebastian Freitag* Dominik Rausch* Yuen C. Law*
Benjamin Weyers* Torsten W. Kuhlen†

*Virtual Reality Group, RWTH Aachen University
JARA – High-Performance Computing
Aachen, Germany
{zielasko, freitag, rausch, law, weyers}@vr.rwth-aachen.de

† Jülich Supercomputing Centre,
Forschungszentrum Jülich
JARA – High Performance Computing
Jülich & Aachen, Germany
t.kuhlen@fz-juelich.de

ABSTRACT

In contrast to the wide-spread use of 6-DOF pointing devices, free-hand user interfaces in Immersive Virtual Environments (IVE) are non-intrusive. However, for gesture interfaces, the definition of trigger signals is challenging. The use of mechanical devices, dedicated trigger gestures, or speech recognition are often used options, but each comes with its own drawbacks. In this paper, we present an alternative approach, which allows to precisely trigger events with a low latency using microphone input. In contrast to speech recognition, the user only blows into the microphone. The audio signature of such blow events can be recognized quickly and precisely. The results of a user study show that the proposed method allows to successfully complete a standard selection task and performs better than expected against a standard interaction device, the Flystick.

Keywords

Interaction techniques; hands-free interaction; non-speech sound interaction; trigger

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Voice I/O*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual reality*; J.5 [**Arts and Humanities**]: Performing Arts

1. INTRODUCTION

In IVEs, the most common interaction mechanism is the usage of a 6-DOF pointing device with mechanical triggers, such as a wand or the ART Flystick. While such devices offer an effective method for selection and manipulation tasks within the IVE, they may prove to be too intrusive. In recent years, advances in vision-based tracking and gesture recognition have allowed gestural interaction methods, where users are not required to wear input devices or tracking markers. However, for gesture interfaces, the definition of trigger signals is challenging. Beside the use of dedicated mechanical devices, such as a thumb switch [7], the use of dedicated trigger gestures [3, 15], e.g., tapping in space [11], is a suitable option in line with a gestural interface. However, gesture recognition is still error-prone and suffers from per-user differences. Furthermore, performing the trigger gesture may interfere with other interaction gestures. Another alternative is speech recognition, which has the drawback of suffering from mis-recognitions and works best only if trained for individual users. For both, gesture and speech recognition, time delay from the recognition algorithm can cause further problems and introduce significant latency into the system [6], which should be avoided especially in IVEs.

The main contribution of this paper is an alternative trigger approach for hands-free interaction scenarios, which allows to precisely trigger events with low latency by blowing into a microphone. We validate the approach by comparing the performance in terms of speed and accuracy against a standard interaction device (Flystick) in a user study.

The rest of the paper is structured as follows. First, we discuss related work in terms of non-verbal vocal input (NVVI) in Section 2. Furthermore, we present our clicking metaphor in Section 3 and evaluate it against a common button trigger in Section 4. Finally, we discuss the results of the user study and point out future work in Section 5.

SUI 2015, August 8–9, 2015, Los Angeles, California, USA.

ACM 978-1-4503-3703-8/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2788940.2788953

2. RELATED WORK

Non-verbal vocal input (NVVI) is a common method to enrich the interaction space of people with physical disabilities, e.g., to steer a wheelchair [4]. However, these techniques are usually very rich, as they use different combinations of voice characteristics, such as pitch and volume, and thus are not ad-hoc accessible to everybody. The *Whistling User Interface* (U^3I) [14, 17], *The Vocal Joystick* [5] and the approach of Chanjaradwichai et al. [2] are recent examples for NVVI interfaces that allow motor-impaired people to use native desktop applications. In case of disabilities, the gain in interaction possibilities usually compansates for the time needed for learning.

Patel and Abowd [13] propose an interface, called *Blowable User Interface* (BLUI), which allows the user to trigger a localized click in a desktop or laptop environment by blowing. They classify the air pressure signatures of the signals recorded by a fixed-positioned microphone and assign them to 1 of 9 cells on the screen. The drawback of this design is that it requires a meaningful and fixed placement of the microphone and a calibration phase for the classifier. Furthermore, a click is not performed until the user first selects a widget by blowing and than blows harder for about one second.

Figure 1: Experimental setup in the CAVE using a standard Fitt's law task (here T3) according to ISO 9241-400:2007.

This brings the latency in regions of speech recognition, which we want to avoid.

Igarashi et al. [8] stated that using non-verbal characteristics of voice also could be beneficial as an interaction technique in virtual environments. Zielasko et al. [19] used an extended blowing metaphor to trigger tones out of bottles in a IVE using an HMD with a gestural interface and the fingers as pointing devices.

3. BLOWCLICK

The idea of BlowClick is to realize a trigger by blowing into a microphone, which may already be part of the setting when also using speech recognition and therefore should neither disturb the latter nor exclude talking to other users. Therefore, the user's breath is captured by a microphone and the current signal frame is condensed to a single strength value. This value is exponentially smoothed over time to reduce jitter. If it lies over a given threshold, the BlowClick's device state is changed to *is triggered*. When it falls below the threshold again, it is changed back to *is not triggered*. To avoid triggers caused by speaking in a normal volume, we measured the values produced by speaking in advance and set the threshold above that. With the used microphone (see below) the threshold for the averaged sample frame was 6,10% of the maximum amplitude.

For the implementation of BlowClick, we used OpenAL as audio framework. Using a buffer size of 1378 samples and a sample rate of 40kHz results in a delay of 30ms, with a neglectable processing time of a few microseconds. This meets the above defined latency requirement, because it is under the recommended threshold of 100ms [12].

4. USER STUDY

For validation, we conducted a quantitative user study to compare the performance of the proposed method in terms of speed and accuracy against a standard interaction device (Flystick). The use case of the proposed method lies outside the scope in which a Flystick is applicable, e.g., in a hands-free scenario, but a comparison with an established interaction device nevertheless is helpful. In advance we formulate the following hypotheses:

H1 It is possible to reasonably solve selection tasks with the BlowClick metaphor.

H2 The standard interaction device outperforms the Blow-Click metaphor with respect to speed and accuracy.

H3 Blowing feels more exhausting then pressing a button as trigger.

4.1 Participants

18 subjects (3 female and 15 male, ages M = 28.9, SD = 5.51) participated in the study. They were unpaid and all had prior experience with IVEs. All of our participants had normal or corrected-to vision. The experiment took about 20 minutes per participant, of which 10 minutes were spent in an IVE (a 5-sided CAVE) and the rest with the introduction and completing the questionnaires. The duration of the experiment was determined in a pre-study revealing that participants get exhausted by holding the Flystick. In the experiment a head-tracked stereoscopic view was provided.

4.2 Design

We used a $3 \times 4 \times 21$ within-subject experimental design (3 device conditions and 4 levels of difficulty each with 21 trials). The following three device conditions appeared in counter-balanced order, following a latin squares design:

CF Flystick button as trigger, Flystick as pointing device

BF Blowing as trigger, Flystick as pointing device

BH Blowing as trigger, hand as pointing device

4.2.1 Apparatus

To track the hand, we use a light weight tracking target by ART that was strapped to the back of the hand. As Flystick, we used the Flystick2 by ART. During the whole experiment the participants wore a wireless microphone, a Sennheiser EW 300 G2 with a Sennheiser ME 3 as actual sensor. The windscreen of the microphone was removed. The threshold used to trigger a click with the microphone was identical for all participants. We added the hybrid device combination BF to the experimental design to be able to examine possible influences of the type of pointing device.

4.2.2 Procedure

For each device condition, 4 Fitts' Law selection tasks with increasing difficulty, designed according to ISO 9241-400:2007 [9, 16] had to be performed (see Figure 1). We measured the time between each successful selection and the total number of clicks performed during the task. Furthermore, we defined an error as a click not leading to a selection, this also includes false positives invoked by sneezing, coughing or screaming, in case of the blowing trigger. The participants were asked to prioritize accuracy over speed. As pointing feedback a simple ray was drawn, starting from the tip of the used device. No extended selection strategy was used. Each task consisted of 21 spheres arranged in a circle with a radius of 0.75m and placed 2.625m in front of the user, with the restriction to hold position in the center of the CAVE. The projection plane of the spheres lay exactly on the CAVE's back wall to exclude any effects of distance estimation [1] and reduce possible effects of target distance, as i.e. reported by [18].

The current target sphere was colored in green (see Figure 1), the currently focused, if any, in blue and all others in white. While a button was pressed, a sphere switched its color from blue to white. The sphere size varied from an easy first task (**T1**) with radius 0.1m over a second task (**T2**) with radius 0.075m and a third task (**T3**) with a radius of 0.05m, to a very difficult last task (**T4**) with a radius of 0.025m. Each of the three device combinations was introduced by a simple training task with 11 spheres of radius 0.15m and no time restriction. Additionally, the participants were told that they can take a break between the different device conditions and between the tasks.

Before the experiment, each participant filled out a demographic questionnaire and was orally briefed about the task. They had no further explanation how to use BlowClick, than "by blowing into

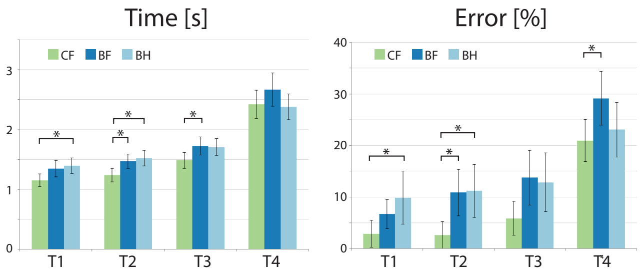

Figure 2: Left: mean time between successful selections in T1-T4, right: clicks not leading to a selection in T1-T4. Error bars show the 95% confidence intervals.

Table 1: P-values and effects for the time and the rate of error, between the different devices.

Task		Time effect [s]	p-value	Error effect [%]	p-value
T1	CF - BF	-.197	.065	-3.823	.388
	CF - BH	-.245	**.015***	-6.999	**.020***
	BF - BH	-.048	1.000	-3.176	.618
T2	CF - BF	-.233	**.018***	-8.241	**.016***
	CF - BH	-.285	**.003***	-8.588	**.011***
	BF - BH	-.052	1.000	-.347	1.000
T3	CF - BF	-.242	**.043***	-7.881	.058
	CF - BH	-.223	.070	-6.990	.110
	BF - BH	.019	1.000	.891	1.000
T4	CF - BF	-.243	.423	-8.206	**.047***
	CF - BH	.043	1.000	-2.126	1.000
	BF - BH	.286	.254	6.080	.208

* .05 level of significance

the microphone". After performing the tasks, the participants were asked to fill out a qualitative questionnaire regarding the subjective usefulness and a device comparison (see Figure 3 and 4), and write down freetext comments.

4.3 Results

We averaged the time between successful selections over all 21 trials for each participant and further averaged these results per task (see Figure 2 left). Furthermore we set the number of errors in relation to the number of total clicks triggered in every task and averaged them over all participants (see Figure 2 right). Table 1 lists p-values and effects for the time and the rate of error, within task difficulties, between the different devices. As expected, the time and error increases with the difficulty of the tasks. However, the data only partially supports hypothesis H2. On the one hand, within the easier tasks T1 and T2, the classic button trigger significantly outperforms both blowing device combinations in 6 out of 8 cases (see Figure 2 and Table 1), regarding time and error, but on the other hand, in the more difficult tasks T3 and T4, only in 2 out of 8 cases. When significantly better, the effect of the standard technique does not show more than 20% increased speed and less than 8% fewer errors. This supports our main hypothesis H1, that it is possible to reasonably solve triggering tasks with the BlowClick metaphor. We did not find any interesting intra-task results.

Figure 3, 4 and Table 2 show the results of the post-study questionnaires. First of all, the results for Q5 also subjectively support H1. Question Q9, Q10 and Q12 show that in case of perceived speed, precision and overall success, the participants had no clear favorite out of standard trigger and blowing, which does not support H2. Question Q2 and Q11 clearly support H3 that blowing is exhausting, even if we do not see any effects over time in the

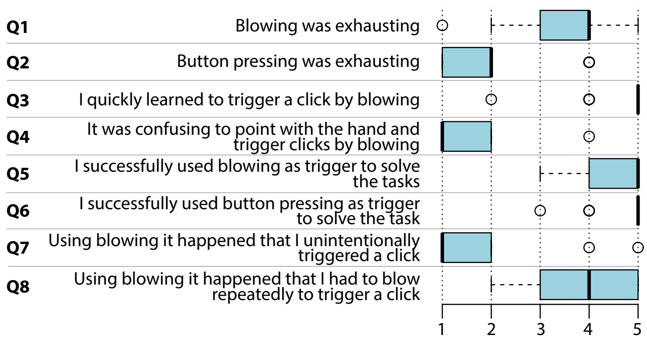

Figure 3: Results of a 5 point Likert scale questionnaire.

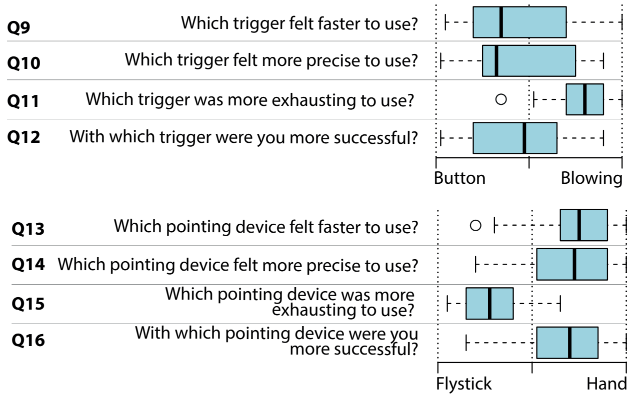

Figure 4: Results of a comparison questionnaire, inspired by NASA TLX.

data. As a spin-off result, it is interesting to notice that participants clearly preferred the hand over the Flystick as pointing device in combination with BlowClick (Q13-Q16).

5. DISCUSSION & FUTURE WORK

The results show that BlowClick as a metaphor for triggering, in selection tasks is a suitable solution. Additionally, it performs better than expected compared to a standard device. In the future, we want to investigate whether these results are repeatable in common application scenarios for IVEs, where the focus does not exclusively lie on the selection method. The results show also that the blowing was perceived as exhausting, which can be a problem in practice. However, the chosen experimental setup, with the goal to blow about a hundred times in a few minutes, does not sufficiently represents all real use cases. Additionally, we observed and

Table 2: Mean (M) and standard derivation (SD) for the questionnaires from Figure 3, scaling discrete from 1 to 5 and Figure 4, scaling from 0 to 20 in .5 steps.

	M	SD		M	SD
Q1	3.67	1.19	**Q9**	8.11	6.22
Q2	2.11	1.13	**Q10**	8.11	5.73
Q3	4.67	0.77	**Q11**	15.14	3.53
Q4	1.56	0.78	**Q12**	8.47	5.68
Q5	4.56	0.62	**Q13**	13.97	4.57
Q6	4.72	0.57	**Q14**	13.39	5.31
Q7	1.61	1.14	**Q15**	5.19	3.83
Q8	3.72	1.13	**Q16**	12.78	5.28

got reported that a part of the participants spent much less effort to trigger a click by blowing than others. They relatively quickly found a way to blow directly into the microphone in a way that needs less effort and especially was possible to perform decoupled from their normal breathing rhythm. However, nearly all participants reported that they were able to quickly learn to trigger a click by blowing (Q3). Thus, it will be interesting to investigate the effects in a longitudinal study. Furthermore, we are convinced that the observed speed and error rate with the BlowClick metaphor can be further reduced. One possibility is to add a visual or auditory feedback that reveals how far away the current amount of blowing is away from triggering a click, w.r.t. the threshold. We are confident that this would reveal the reason why an intended click did not happen and additionally could give confidence to the user that a click did not happen not because of the blowing, but the pointing. We sometimes observed that participants increased the amount of blowing more and more when a series of errors happened, when the reason for that actually was not the blowing. Second, the underlying framework for the study triggered a click event on a sphere only when the *trigger down* and *trigger up* event both happened while focusing the sphere. While this is a valid method to evaluate a click, some participants reported that this was confusing to them or even led to a lot of errors, because they already aimed for the next target and only then noticed that the last sphere had not been selected. This fact influenced both trigger methods, but the influence should have been stronger with the blowing, as its duration was normally longer and so the probability that the sphere was already left was higher. Additionally, Isokoski [10] noted that there are significant differences in the performance of a computer mouse just with respect to the actual button event evaluated as trigger. Possible solutions for further investigations are, to better prepare the participants, give a clearer visual or auditory feedback that a click was performed, or that it just disappears due to the learning in a longitudinal study. Finally we want to improve the blow detection by trying to even better decide if the current audio signal originates from speech or blowing, e.g., by considering the amount of signal clipping. This would further increase the usability in many use cases.

6. CONCLUSION

In this paper, we presented a low-latency approach to precisely trigger events by blowing into a microphone. The results of the performed user study show that the proposed method allows to complete a standard selection task and performs better than expected against a standard interaction device.

7. REFERENCES

[1] G. Bruder, F. A. Sanz, A.-H. Olivier, and A. Lécuyer. Distance Estimation in Large Immersive Projection Systems, Revisited. *In Proc. of IEEE Virtual Reality*, pages 27–32, 2015.

[2] S. Chanjaradwichai, P. Punyabukkana, and A. Suchato. Design and Evaluation of a Non-Verbal Voice-Controlled Cursor for Point-And-Click Tasks. *In Proc. of International Convention on Rehabilitation Engineering & Assistive Technology*, pages 48:1–48:4, 2010.

[3] A. Choumane, G. Casiez, and L. Grisoni. Buttonless Clicking: Intuitive Select and Pick-Release Through Gesture Analysis. *In Proc. of IEEE Virtual Reality*, pages 67–70, 2010.

[4] L. Fehr, W. E. Langbein, and S. B. Skaar. Adequacy of Power Wheelchair Control Interfaces for Persons with Severe Disabilities: A Clinical Survey. *Journal of Rehabilitation Research and Development*, 37(3):353–360, 2000.

[5] S. Harada, J. Landay, and J. Malkin. The Vocal Joystick: Evaluation of Voice-Based Cursor Control Techniques. *In Proc. of International ACM SIGACCESS Conference on Computers and Accessibility*, pages 197–204, 2006.

[6] S. Harada, J. O. Wobbrock, and J. A. Landay. Voice Games: Investigation Into the Use of Non-speech Voice Input for Making Computer Games More Accessible. *In Proc. of IFIP International Conference on Human-Computer Interaction*, pages 11–29, 2011.

[7] B. Hentschel, J. Künne, I. Assenmacher, and T. Kuhlen. Evaluation of a Hands-Free 3D Interaction Device for Virtual Environments. *Virtuelle und Erweiterte Realität: Workshop der GI-Fachgruppe VR/AR*, pages 149–156, 2015.

[8] T. Igarashi and J. F. Hughes. Voice as Sound: Using Non-Verbal Voice Input for Interactive Control. *In Proc. of ACM Symposium on User Interface Software and Technology*, 3(2):155–156, 2001.

[9] ISO. *Ergonomics of Human-system Interaction: Principles and requirements for physical input devices (ISO 9241-400:2007, IDT)*. International Organisation for Standardisation, 2007.

[10] P. Isokoski. Variability of Throughput in Pointing Device Tests: Button- Up or Button-Down? *In Proc. of ACM NordiCHI*, pages 68–77, 2006.

[11] Y. Jang, S.-T. Noh, H. J. Chang, T.-K. Kim, and W. Woo. 3D Finger CAPE: Clicking Action and Position Estimation under Self-Occlusions in Egocentric Viewpoint. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):501–510, 2015.

[12] J. Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., 1993.

[13] S. N. Patel and G. D. Abowd. BLUI: Low-Cost Localized Blowable User Interfaces. *In Proc. of ACM Symposium on User Interface Software and Technology*, pages 217–220, 2007.

[14] O. Poláček, A. J. Sporka, and P. Slavík. A Comparative Study of Pitch-Based Gestures in Nonverbal Vocal Interaction. *IEEE Transactions on Systems, Man and Cybernetics*, 42(6):1567–1571, 2012.

[15] J. Segen and S. Kumar. Shadow Gestures: 3D Hand Pose Estimation Using a Single Camera. *In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:479–485, 1999.

[16] R. W. Soukoreff and I. S. MacKenzie. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. *International Journal of Human Computer Studies*, 61(6):751–789, 2004.

[17] A. J. Sporka, S. H. Kurniawan, and P. Slavik. Whistling User Interface (U^3I). *ERCIM Workshop on User Interfaces for All*, 3196:472–478, 2004.

[18] R. J. Teather and W. Stuerzlinger. Pointing at 3D Target Projections with One-eyed and Stereo Cursors. *Proc. ACM CHI*, pages 159–168, 2013.

[19] D. Zielasko, D. Rausch, Y. C. Law, T. C. Knott, S. Pick, S. Porsche, J. Herber, J. Hummel, and T. W. Kuhlen. Cirque des Bouteilles: The Art of Blowing on Bottles. *In Proc. of IEEE Symposium on 3D User Interfaces*, pages 209–210, 2015.