2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation

Leveraging Blowing as a Directly Controlled Interface

Yeqing Chen[1], Yulong Bian[1]*, Chenglei Yang[1, 3], Xiyu Bao[1], Yafang Wang[1],
Gerard de Melo[2], Juan Liu[1], Wei Gai[1], Lu Wang[1] and Xiangxu Meng[1, 3]
[1]*School of Computer Science and Technology*
Shandong University, Jinan, China
Email: 2931852768@qq.com
[2]*Rutgers University, New Brunswick, USA*
[3]*Engineering Research Center of Digital Media Technology, MOE, Jinan, China*

***Abstract*—Breathing is a natural and directly controllable human interaction method. Currently, only a few works have considered breath as a direct input controlling mechanism. The equipment relied upon in these works is generally complicated, expensive, inconvenient to wear, and sometimes insufficiently controllable. This paper proposes a simple method to implement a natural and directly controllable interaction based on blowing air. The method can provide interaction operations to applications by transferring blowing of air into sound using *headset microphones* and classifying them into different categories of blowing air in conjunction with support vector machines and particle swarm optimization. During tests, it's proved that this interface not only increases the type of blowing used for interaction, but also eliminates interference from speaking in a normal volume effectively. This user interface can be conveniently used in special environments (e.g., noise, unavailability of hands) or for special groups (e.g., deaf-mutes).**

***Keywords*-natural interaction; blowing; microphone; noise elimination**

I. INTRODUCTION

From classic interaction modalities such as keyboard and mouse to natural interactions such as multi-touch, voice, gesture and posture, eye tracking and brain-computer interaction, the field of human-computer interaction technology has made substantial progress. However, these forms of interactions are not necessarily suitably in all scenarios, e.g. when one's hands are unavailable for mouse or touch interaction, in noisy or speechless environments for voice interaction, or for physically challenged people in the case of speech or eye tracking interaction. Recently, breathing has been considered as an alternative control mechanism to influence the physical world and the virtual environment [16]. The convenience and controllable nature of breathing interactions can sometimes make up for the disadvantages of more common interactive modes. Thus far, some works have studied breathing or blowing air as a direct input modality in proper detail. However, the equipment (e.g., breathing sensors, breathing belts, or other special sensors) used in these works is generally complicated, expensive, inconvenient to wear, and on occasion insufficiently controllable, while also lacking universal portability in daily life [11], [16].

In this paper, we propose a simple method to implement a natural and directly controllable interaction based on blowing air. We put together a sensor setup(*headset microphones*) lighter than previous work, a different sensing algorithm, and did our evaluation embracing the environmental noise and speaking noise, which is the main delta against previous literature. We designed three application examples (i.e., playing video on a PC, accessing web pages on a mobile phone, and playing a VR game) to assess the merits of our method. The experiments and applications show that our method is effective, controllable, and convenient to use as a natural form of interaction. This interface can also be suitable for special environments (e.g., noise, unavailability of hands) or for special groups (e.g., deaf-mutes).

II. RELATED WORK

Breathing is a human instinct. Because breathing can be consciously controlled, some work uses it as a directly interactive method. This section summarizes some of the most directly related work on interaction based on breathing or blowing air.

Some researchers used wearable sensors or custom sensors to obtain physiological signals(e.g., the amount and speed of the exhaled air, the piezoelectric signal as well as the temperature) of breathing for some direct mapping operations or interactive control. There are some studies [1], [6], [8], [13], [14], [16] using wearable breathing sensors, which are inconvenient and expensive to use, and not common in everyday life. Those specially-customized devices in these studies [3], [5], [9], [11], [12], [15], [17] are also inconvenient and expensive to use, and have great limitations in terms of usage scenarios and usage methods. All of the above work generally either relies on custom devices such as breathing sensors, breathing belts, or other special sensors, which are intricate, expensive, inconvenient to wear, lack universal portability in daily life, and may sometimes be less controllable, which has great limitations in the usage mode and scenario and has not been further explored.

There is a low-cost way of interacting with breathing, using a microphone to obtain the sound of breathing for interactive control. Blowing into the microphone has been a popular input method for smartphones games and music

978-1-7281-4034-6/19/$31.00 ©2019 IEEE
DOI 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00115

applications since the Ocarina by Wang [19], [20]. Misra et al. explored to use microphone as a generic sensor in MobileSTK to drive sound synthesis algorithms in expressive ways [7]. Patel and Abowd presented a coarse-grained system, called BLUI, that enables blowing at a laptop or computer screen to directly control interactive applications [10]. Zielasko et al. presented an alternative trigger approach for hands-free interaction scenarios to precisely trigger events by blowing into a microphone [21]. When the blowing value exceeds a given threshold, event is triggered, otherwise, the event is not triggered. Filho et al. proposed a mobile phone interface by exploring the processing of the audio from the microphone in mobile phones to trigger and launch software events [2]. Because mobile phones only have simpler computing processes due to limited processing power, compared to laptops and desktops, it is difficult to handle complex types of blowing operations, like identifying blowing sound and speaking voice.

Our research in this paper instead only needs regular microphones to obtain data pertaining to the exhaled airow, which is effective, controllable, and convenient. This paper classifies blowing sound into different categories as directly controlled interactions using machine learning method, which not only increases the type of blowing used for interaction, but also avoids triggers caused by speaking in a normal volume effectively.

III. SYSTEM ARCHITECTURE

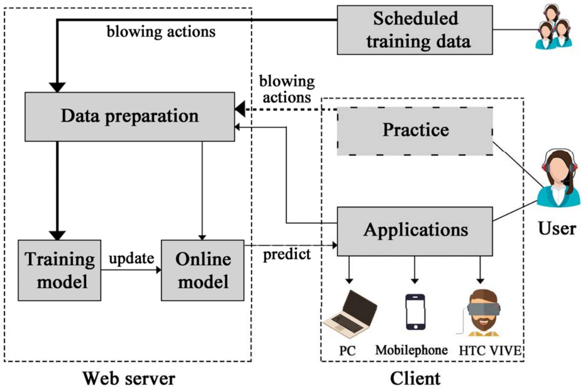
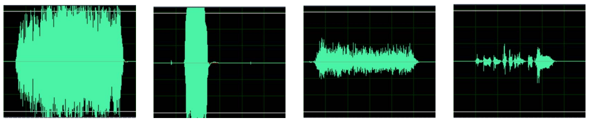


Figure 1. System Architecture.

Fig. 1 provides an overview of the system architecture. A web server works as a back-end for processing data, training, and running models. When a person uses the interaction interface, the prediction model running on the server obtains that person's breath data transmitted from the client to recognize air blowing actions, and sends these back to the client (e.g., PC, mobile phone, or HTC VIVE). Once the clients receive the action information, the running application will perform the corresponding operations.

To improve the performance of the recognition, we ordinarily update the model by collecting new training data in two ways. It will describe in detail in *Training data acquisition* part.



(a) *gale* (b) *gust* (c) *breeze* (d) *noise*

Figure 2. Sound waveforms obtained by the microphone for three types of blowing air and two examples of speaking voice.

Air Blowing Interaction Design Scheme. Referring to the work of Sra et al. [16], we distinguish four forms of blowing air with regard to duration and intensity. Among them, three are used for possible interactions(cf. Figs. 2) and one is used to eliminate interference from speaking voice(one example in the Fig. 2(d) belong to the category of speaking voice):

gale : very strong exhaling sustained for 2-3 seconds (cf. Fig. 2(a) and Fig. 3(a)).
gust : strong jet but transient for a short duration of less 1 second (cf. Fig. 2(b) and Fig. 3(b)).
breeze : slow and gentle blowing for 2-3 seconds (cf. Fig. 2(c) and Fig. 3(c)).
noise : speaking voice in a normal volume. (cf. Fig. 2(d)).

One important consideration is the placement of the microphone. An effective position is near the mouth and pointed towards the mouth(cf. Fig. 5 and Fig. 6). Figs. 2 shows that the waveform of blowing interactions and speaking voice are very different from each other, so our blowing interaction will be still effective even if the user speaks. Our experiments also verified it.

IV. THEORY AND MODEL

The proposed interface acquires the sound signals resulting from blowing air in real time and then determines the particular form of blowing air as interactions for different applications.

A. Training data acquisition

The training data can be collected in two ways:

1) Scheduled training data: To collect a sufficiently large-scale dataset for training at the very beginning, we rely on audio recording with a sampling frequency of 8000 Hz. In order to make the initial training set more standardized, participants are required to blow air every 3 seconds. In the experiments, we relied on two participants for training data collection.

2) User practice: Due to the different intensity levels when blowing air, users are able to practice before using the interface so as to obtain better results. During practice, every user is instructed to perform the correct kind of exhalation in accordance with the indicated form of blowing air. Every user's personal data is uploaded into the server's training set. And before this user use our interface, our prediction model will be trained using the updated training set. This is not compulsory, and users can use our interface directly without first practicing.

B. Data preparation

1) Signal preprocessing: The sound signal is first normalized to a standardized range of $[-1, 1]$. Then a sliding window is applied for sampling discretization. The continuous signal in this normalized data is then divided into discrete segments using a sliding window of 24000 Hz with 67% overlap between segments [4].

2) Feature extraction: The classifier operates based on four signals extracted as features from the discrete sound data. These are the mean (given by (1)), variance (given by (2)), the proportion of the number of normalized values greater than 0.4 (given by (3)), and 0.6 (given by (4)).

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{1}$$

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \tag{2}$$

$$p1 = \frac{|\{X_i > 0.4\}|}{n} \tag{3}$$

$$p2 = \frac{|\{X_i > 0.6\}|}{n} \tag{4}$$

where X_i is the i-th element in a discrete sample, and n is the length of each sample. The feature $p1$ and $p2$ is chosen to account for both the strength and duration characteristics of the exhalation.

C. Training and prediction

1) Classifier: In our system, we rely on support vector machines (SVMs) to identify the category based on the features. Although other learning algorithms are applicable as well, SVMs are powerful tools used to solve the small sample, nonlinear relationships and multiple classification problems [18]. Their ultimate classification performance depends heavily upon the selection of penalty parameters and appropriate kernel.

2) Training: In the initial training process of the classifier (no user practice data is added to the training set), we use a particle swarm optimization (PSO) algorithm to optimize the penalty parameter c in the SVM and the parameter g in the RBF kernel function to improve the classification accuracy.

Given the training data acquired using the two procurement schemes described above, the classifier is trained using training data preprocessed as described above. The trained model is periodically updated as new data comes in.

Algorithm 1 Air blow category recognition

```
Input: $\theta$ is the threshold;
while our interface is being used do
    get real-time volume $v$ of blowing sound;
    if $v > \theta$ then
        get sound data $d$;
        $d'$=prepare($d$);
        $a$ = recognize($d'$);
        client performs interactive operations
            according to $a$;
return
```

3) Prediction: Algorithm 1 provides the details of the blowing type recognition.

V. USER EXPERIMENTS

To assess the usability and effectiveness of our interaction method leveraging exhalation actions, we conducted a user study.

A. Participants

A total of 16 student volunteers (6 females, 10 males) were enrolled to participate in the study. The age of the sample ranges from 14 to 24 years (M = 16.94 years, SD = 3.83 years). Before this test, all participants did not have any prior experience of blowing air as an interaction method.

B. Application design

Our novel blowing-based interface can be applied in a range of different applications.

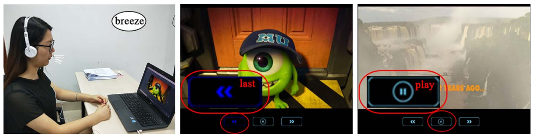
1) Playing video on a PC: The client relies on Unity 3D (version 5.6.0) as the platform for showing video. It obtains the exhalation sound waves in real-time, and monitors which blowing operation the user has blown in real-time. When monitoring any blowing actions, it performs operations corresponding to the identification and classification result: *gale* (cf. Fig. 3(a)), *gust* (cf. Fig. 3(b)), *breeze*(cf. Fig. 3(c)).

2) Accessing web pages on a mobile phone: We created a simple prototype of an interactive web page on a mobile phone, for which the action of entering a selected link, as can be done via the Enter key, is mapped to the *gust* action (cf. Fig. 4), while the sequential jumping between successive input controls, as can be achieved with the Tab key, is mapped to the *breeze* action (cf. Fig. 5).

3) Playing a VR game: We integrate the blowing actions into a VR game called Undersea Treasure Hunt. It is developed using Unity 3D (version 5.6.0), and played using an HTC VIVE headset. The three game effects are associated with blowing actions: *gale* sprays a water jet, *gust* will open a treasure box, *breeze* triggers the bubbling operation of the crab (cf. Fig. 6).

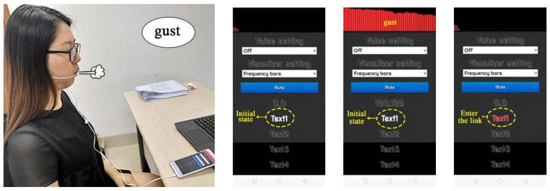
(a) *gale* controls the play or pause action

(b) *gust* for the next video



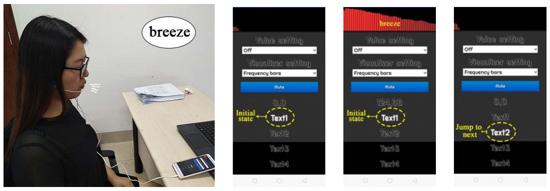
(c) *breeze* for the previous video

Figure 3. Blowing actions and corresponding effects on a PC.



(a) *gust* (b) Initial state (c) *gust* (d) Enter the link

Figure 4. *Gust* corresponds to enter.



(a) *breeze* (b) Initial state (c) *breeze* (d) Jump to next

Figure 5. *Breeze* corresponds to tab.

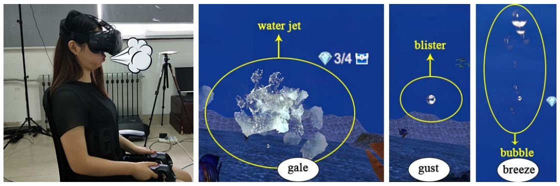


Figure 6. Blowing actions and corresponding effects in a VR game.

C. Experimental design

The user study consisted of two parts: a usability test and user experience study.

1) Usability test: We adopted a within-subjects design for the usability test. The first application, the PC application, was used for this study. To assess the usability of the interaction technology adequately, we tested the interaction accuracy in five conditions: a quiet environment vs. four noisy environments. In noisy environments, music was played at different distances with different volumes(65db with 0.5 meters, 65db with 0.1 meters, 80db with 0.5 meters and 80db with 0.1 meters), serving as noise for the test. The experience order was counter-balanced.

2) User experience study: In this second part, we used the web application and the VR game application to explore the participant's experience of using the interaction method. After the participant experienced the two applications, we conducted a follow-up interview. The interview included four structural elements: (1) interest; (2) applicability; (3) generalization and (4) disadvantages. Moreover, we added an open topic item regarding blowing air as a natural and directly controllable interaction method.

D. Procedure

First, we designed a practice module, and the participant practiced the three types of blowing following a simple instruction phase before starting the formal test.

Then, we explained to the participants the relevant interaction tasks relevant to the test. The participants experienced the application in five environment conditions using three blowing interactions. In every environment the test consisted of nine rounds(each round included 6 interaction tasks: two times for each type of the three blowing types) and 6 times of participant speaking, i.e., 60 times of blowing. The accuracy of the actual interaction would be recorded. To avoid the participants from developing a certain regularity, the sequence of six breathing tests in each round was random.

After completion of Test 1, the participants experienced the web application and VR game application, respectively. The participants were able to take breaks during the test. After experiencing the two applications, we requested a follow-up interview with the participants.

E. Results

1) Accuracy: The results of the accuracy in the usability test for the sixteen participants are given in Table I.

Table I
THE AVERAGE ACCURACY OF 16 PARTICIPANTS IN FIVE ENVIRONMENTS

Environment	Average Accuracy
Quiet	90.52%
Noisy 1: 65db+0.5m	89.17%
Noisy 2: 65db+0.1m	86.83%
Noisy 3: 80db+0.5m	88%
Noisy 4: 80db+0.1m	81.33%

From the test results, we can observe that our average accuracy is 96.3% in cross-validation and a bit lower in the usability test (Table I), but still comparable with the previous research [16] (the average recognition accuracy in this study was 88.3% with two authors) that required custom hardware. Therefore, the accuracy of interaction technique in this study is deemed acceptable.

Moreover, compared with the quiet environment, the accuracy rate did not decrease much in Noisy 1, Noisy 2, Noisy 3, while dropped in Noisy 4. Noisy 4 (80db and 0.1 meters) was a noisy environment by interviewing participants. It showed that our technology can cope well with potential interference stemming from common background noise.

2) Interview results for Test 2:

Interest. Through experience with the web application and VR game application, the participants felt that this interaction method is very interesting, e.g. "*I have interacted with the mouse, touch, and voice, but not breath, and it was funny*".

Applicability. It may solve problems under certain special circumstances or for people with particular needs. "*In some special dangerous situations or in special emergencies, such as when a hostage needs to call for help, he can't shout, call or send messages. By using this interaction, he can send a secret signal for help, which is not easily detected by criminals*".

Generalization. Although it cannot be regarded as a primary way of interaction, it can be used as an auxiliary interaction modality. In some cases, it may replace touching and other interaction forms. "*In some cases it can replace voice and touch, for example, when both hands are inconvenient to use, you can blow to answer the phone*".

Disadvantages. First, the number of interaction types that can be achieved is limited. "*Unlike touch and voice, which can achieve many kinds of control*". Second, it easily leads to fatigue. "*Although it is an interesting interaction method, the user will be tired after using many times*".

VI. DISCUSSION

A. Special advantages

1) The interaction is natural and convenient: Blowing is a natural and directly controllable interaction method.

2) The device is the simplest and most commonly used in daily life: The only device used in our blowing interaction setup to obtain the needed data are *headset microphones*, which are not only common, simple, and cheap, but also convenient to carry around in dairy life.

B. Interaction accuracy

From the accuracy in the current test, participants speech and the environmental noise had little influence on our interaction based on blowing out, but the proficiency of the interaction types had a great influence on the accuracy rate. If our blowing interactions are applied to common operations, the accuracy rate will certainly increase with the increase of proficiency in using blowing interaction.

There are three main reasons for this: (1) In the actual application process, there will be many uncontrollable factors. (2) The classification model can be further optimized. (3) Only two people's data was used for the initial training and cross-validation. Hence, there is also a need for further research on improving the algorithms.

C. Applicability

1) Assisting other interaction modes: The incorporation of blowing interactions into other setups can liberate the hands. Example settings include answering the phone while driving, zooming in and out during navigation, learning to cook in a noisy kitchen, and so on.

2) Improving the user experience in using applications: We observed that a user's interest and experience is greatly enhanced if this interaction is designed in an appropriate context, such as a VR game.

3) Generalization to special groups and special situations: This interaction modality has a high degree of availability in special environments (e.g., noisy ones, unavailability of hands) or for special groups (e.g., deaf-mutes). This approach may prove useful in certain emergency situations.

D. Limitations and future work

First, the number of interaction types that can be achieved is limited. Since memory space is limited and there are not that many forms of blowing that people can skillfully achieve, there are not many categories of blowing interactions that can reliably be supported, which is a limitation. Second, in very noisy cases, the interaction can be disturbed. Thus, we plan to further optimize the algorithm to improve the recognition accuracy.

VII. CONCLUSION

This paper presented a simple method to implement a natural and directly controllable interaction based on blowing out air. We designed and implemented three blowing actions and eliminated interference from environmental noise as well as speaking in a normal volume effectively, only relying on *headset microphones* to record the audio signal and using optimized classify it to distinguish different blowing categories.

We found that environmental noise had little influence on the interaction and that this interaction modality is a good way to improve users' interest and experience in applications, specifically in VR applications. Moreover, it is not only very available in a multitude of environments, but also particularly usable in special environments (e.g., noise, unavailability of hands) or for special groups (e.g., deaf-mutes).

ACKNOWLEDGMENT

This work is supported by the National Key Research and Development Program of China (2018YFC0831003), Shandong Key Research and Development Program

(2017CXGC0606), and the National Natural Science Foundation of China (61802232).

REFERENCES

[1] K. Höök, M. P. Jonsson, A. Ståhl, and J. Mercurio, “Somaesthetic appreciation design,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16, 2016, pp. 3131–3142.

[2] F. F. Jackson, W. Prata, and T. Valle, “Breath mobile: a low-cost software-based breathing controlled mobile phone interface.” in *International Conference on Human-computer Interaction with Mobile Devices Services Companion*, ser. MobileHCI’12. San Francisco, CA, USA: ACM, 2012.

[3] D. Jain, M. Sra, J. Guo, R. Marques, R. Wu, R. Chiu, and C. Schmandt, “Immersive terrestrial scuba diving using virtual reality,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’18. New York, NY, USA: ACM, 2016, pp. 1563–1569. [Online]. Available: http://dx.doi.org/10.1145/2851581.2892503

[4] M. M. Javaid, M. A. Yousaf, Q. Z. Sheikh, M. M. Awais, S. Saleem, and M. Khalid, “Real-time eeg-based human emotion recognition,” in *Neural Information Processing*, S. Arik, T. Huang, W. K. Lai, and Q. Liu, Eds. Cham: Springer International Publishing, 2015, pp. 182–190.

[5] K. Kuzume, “Input device for disabled persons using expiration and tooth-touch sound signals,” in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC ’10, 2010, pp. 1159–1164.

[6] J. Marshall, D. Rowland, S. Rennick Egglestone, S. Benford, B. Walker, and D. McAuley, “Breath control of amusement rides,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’11, 2011, pp. 73–82.

[7] A. Misra, G. Essl, and M. Rohs, “Microphone as sensor in mobile phone performance,” in *New Interfaces for Musical Expression*, ser. NIME08, Genova, Italy, 2008, pp. 185–188.

[8] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, “Biofeedback game design: Using direct and indirect physiological control to enhance game interaction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’11. New York, NY, USA: ACM, 2011, pp. 103–112. [Online]. Available: http://doi.acm.org/10.1145/1978942.1978958

[9] Y. Okuno, H. Kakuta, T. Takayama, and K. Asai, “Jellyfish party: Blowing soap bubbles in mixed reality space,” in *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR ’03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 358–359.

[10] S. N. Patel and G. D. Abowd, “Blui: low-cost localized blowable user interfaces.” in *Acm Symposium on User Interface Software Technology*, ser. UIST07. Newport, Rhode Island, USA: ACM, 2007.

[11] R. Patibanda, F. F. Mueller, M. Leskovsek, and J. Duckworth, “Life tree: Understanding the design of breathing exercise games,” in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY ’17, 2017, pp. 19–31.

[12] E. Sawada, S. Ida, T. Awaji, K. Morishita, T. Aruga, R. Takeichi, T. Fujii, H. Kimura, T. Nakamura, M. Furukawa, N. Shimizu, T. Tokiwa, H. Nii, M. Sugimoto, and M. Inami, “Byu-byu-view: A wind communication interface,” in *ACM SIGGRAPH 2007 Emerging Technologies*, ser. SIGGRAPH ’07, 2007.

[13] H. Schnädelbach, K. Glover, and A. A. Irune, “Exobuilding: breathing life into architecture,” in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction 2010, Reykjavik, Iceland, October 16-20, 2010*, 2010, pp. 442–451.

[14] H. Schnädelbach, A. A. Irune, D. S. Kirk, K. Glover, and P. Brundell, “Exobuilding: Physiologically driven adaptive architecture,” *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 4, pp. 25:1–25:22, 2012.

[15] T. Sonne and M. M. Jensen, “Chillfish: A respiration game for children with adhd,” in *Proceedings of the TEI ’16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI ’16, 2016, pp. 271–278.

[16] M. Sra, X. Xu, and P. Maes, “Breathvr: Leveraging breathing as a directly controlled interface for virtual reality games,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’18. New York, NY, USA: ACM, 2018, pp. 340:1–340:12. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173914

[17] P. Tennent, D. Rowland, J. Marshall, S. R. Egglestone, A. Harrison, Z. Jaime, B. Walker, and S. Benford, “Breathalising games: Understanding the potential of breath control in game interfaces,” in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ser. ACE ’11, 2011, pp. 58:1–58:8.

[18] V. N. Vapnik, “The nature of statistical learning theory,” in *Transaction of Beijing Institute of Technology*, 1995.

[19] G. Wang, “Designing smule’s ocarina: The iphone’s magic flute,” in *New Interfaces for Musical Expression*, ser. NIME09, Pittsburgh, PA, 2009.

[20] ——, “Ocarina: Designing the iphone’s magic flute,” ser. 38.2, 2014, pp. 8–21.

[21] D. Zielasko, S. Freitag, and D. Rausch, “Blowclick: A nonverbal vocal input metaphor for clicking.” in *Acm Symposium on Spatial User Interaction*, ser. SUI 2015. Los Angeles, California, USA: ACM, 2015.