

# Vocal Control of a Radio-controlled Car

Adam J. Sporka<sup>1,2</sup>, Pavel Slavík<sup>2</sup>

<sup>1</sup>University of Trento, Department of Information Engineering and Computer Science

<sup>2</sup>Czech Technical University in Prague, Department of Computer Graphics and Interaction

adam.sporka@disi.unitn.it, slavik@fel.cvut.cz

## Introduction

Increasing participation of the community of computer users with special needs in computer entertainment brings about the need for understanding the benefits and drawbacks of different assistive input techniques and devices that can be used for controlling the entertainment experience. Especially challenging is the support of the control of the arcade games, such as *Tetris* [7] or *Breakout* [2], where while only a limited number of game commands is required, the users must issue these in a rapid response to the evolving gameplay – and many assistive devices are unable to cope with this problem. A promising set of techniques is provided by employing the voice as an input device.

Generally speaking, the vocal input is based on engaging the user's voice into the input pipeline of a system. Typically, the voice is registered by a microphone, digitized, and converted into a discrete signal. The signal is being analyzed in an application-dependent way. A traditional instance of vocal input is the automatic speech recognition, where the voice signal is analyzed for words contained in it. The recently introduced non-speech input is based on analysis of the user-produced sounds like humming or whistling [6]. The non-speech input has been reported for example also in [1], [4], and [5].

One particular type of the non-speech input is the input by pitch of tone. It is based on the analysis of the development of the pitch of a tone produced by the user, regardless on the type of the sound the user chooses to use, i.e. whistling, or humming. The input by pitch offers various benefits, including language independence and a very low computational cost.

## The Study

In our previous study [11] we have demonstrated how the user's voice can be used to control the game of *Tetris*.

In the present study we compare the usability of the speech recognition and the non-speech sound input for a direct control of a simple model of a car. It is another instance of a system in which the real-time control of the input is needed. The performance of both methods was compared on a series of three simple steering tasks. The car is a simple radio-controlled model with a Bluetooth interface and a dedicated API, shown in Fig. 1. Two user interfaces have been created: One with the speech recognition, and one using the non-speech sound input.

## Procedure

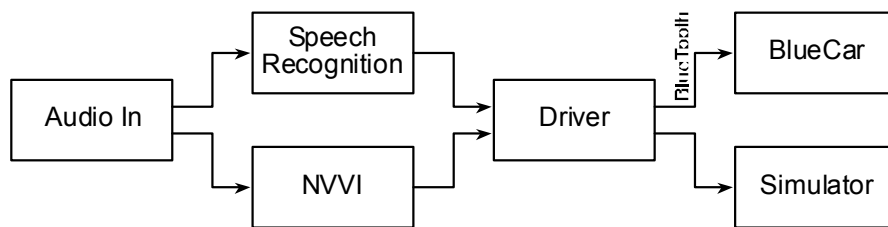
**Participants.** 7 people (6 M, 1 F; 26 years old, SD = 1.61) took part in the study. They all were students of the CTU in Prague. Five of them had a previous experience using both speech recognition technology and the non-speech sound input, gained during the *Tetris* study [11].

**Aparatus.** The BlueCar [3] was used in this experiment. The BlueCar is a mobile robot with Bluetooth interface. The architecture of the experiment set-up is shown in the figure 2.

Depending on which method is being used at the moment, the input audio signal is routed either to the non-speech sound input or speech recognition interface. The speech recognition interface has been based on one of the industry standards, the Microsoft Speech API. The pitch detection has been based on a simple autocorrelation technique. Both interfaces are supposed to recognize commands in the user's input. The commands are then directed to the BlueCar device via Bluetooth connection. The authors of BlueCar also provide a simulator tool which is supposed to demonstrate the function of BlueCar. In our set-up we decided to ask our participants to use the simulator tool as a training platform.

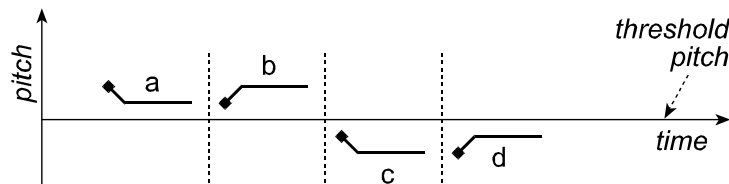


**Fig. 1: BlueCar; from [12]**



**Fig. 2: Voice Control of BlueCar -- Prototype Architecture.**

The commands supported by the speech recognition were „forward“, „backward“, „stop“, „left“, and „right“. The non-speech sound input gestures are displayed in Fig. 3, and cover the equal range of functions. To enable simultaneous control of the direction of movement (forwards or backwards) and the steering, we decided to use a similar approach to the mouse cursor control, as described in [12]. For BlueCar control, the initial pitch of the tone determined whether the car would move forward or backward, depending whether the pitch was above or below a user-specified threshold respectively.



**Fig. 3: Non-speech gestures for the BlueCar control: a ... forward left; b ... forward right; c ... backward left; d ... backward right.**

**Procedure.** Since the performance of using the acoustic input increases with the user's experience, we decided to train the users further before commencing any measurement. According to the results of our previous experiment described in [8] we decided to ask the participants to train using the interface for 5 days, which was therein reported length of the training period. The structure of individual participant's involvement was as follows:

- Day 1: Initial training in the laboratory. The experimenters have demonstrated the function of the user interface and gave instructions on how to download and install the training software.
- Days 2—4: Individual unsupervised training at home. The participants were only asked to report how long they trained each day.
- Day 5: The main measurement in the laboratory, filling out a questionnaire. On the fifth day, the participants were invited to the laboratory again to perform the steering tasks using both methods and fill out a simple post-test evaluation questionnaire. The actual steering tasks were to drive the car through tracks built out of small wooden slabs. Their layout is shown in Fig. 4.
  - Task 1 ... Go 2m forward, then back off to the starting point.
  - Task 2 ... Go 2m forward, take a turn, then return to the starting point.
  - Task 3 ... Drive a curved route.

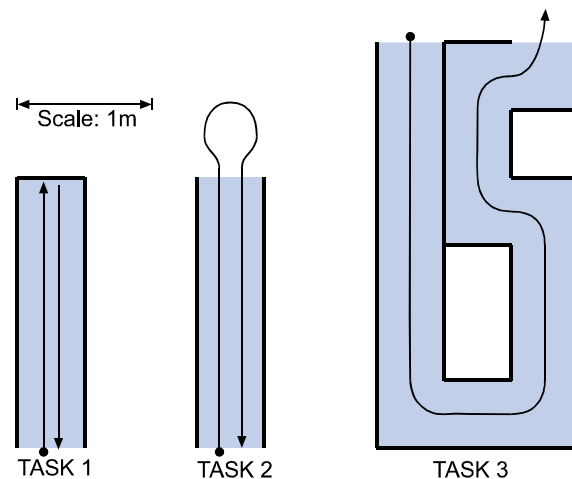


Fig. 4: Steering Tasks

Two values were measured on each performed task: Time needed to complete the task, and the number of penalty points. One penalty point would be awarded for each wooden slab that the car would have pushed over during the course of the task. Due to the failure of the BlueCar's battery, we were not able to record the performance of U7 using the speech recognition.

## Discussion and Conclusion

Table 1 shows the performance of individual users U1 through U7 in each task with both methods. Due to the time constraints, the tasks were not repeated. Table 2 shows the aggregate values and their standard deviations.

**Table 1: Performance Data.**

User	Task 1				Task 2				Task 3			
	Speech		NS		Speech		NS		Speech		NS	
	PP	Time	PP	Time	PP	Time	PP	Time	PP	Time	PP	Time
U1	21	50	0	20	5	80	0	25	29	132	0	17
U2	0	22	15	49	18	70	0	23	26	95	2	18
U3	8	25	0	20	15	60	4	72	22	78	5	40
U4	7	50	0	18	23	72	8	90	16	45	1	20
U5	0	20	0	22	17	63	0	52	22	44	0	37
U6	11	88	0	14	15	89	0	102	29	130	1	57
U7	n/a	n/a	0	21	n/a	n/a	2	53	n/a	n/a	0	68

**Legend: PP ... penalty points. Time is in seconds.**

**Table 2: Aggregate values for each user and all users.**

### Total Time [s]

	U1	U2	U3	U4	U5	U6	Average	SD
Speech	262	187	163	167	127	307	202.2	68.1
NS	62	90	132	128	111	173	116	38.12

### Total Penalty points [-]

	U1	U2	U3	U4	U5	U6	Average	SD
Speech	55	44	45	46	39	55	47.3	6.4
NS	0	17	9	9	0	1	6	6.9

Figure 5 shows the average users' responses reported in the questionnaire. The questionnaires used the 1...5 scale where 1 corresponded to „absolutely false“ and 5 to „absolutely true“ on each particular aspect of the system.

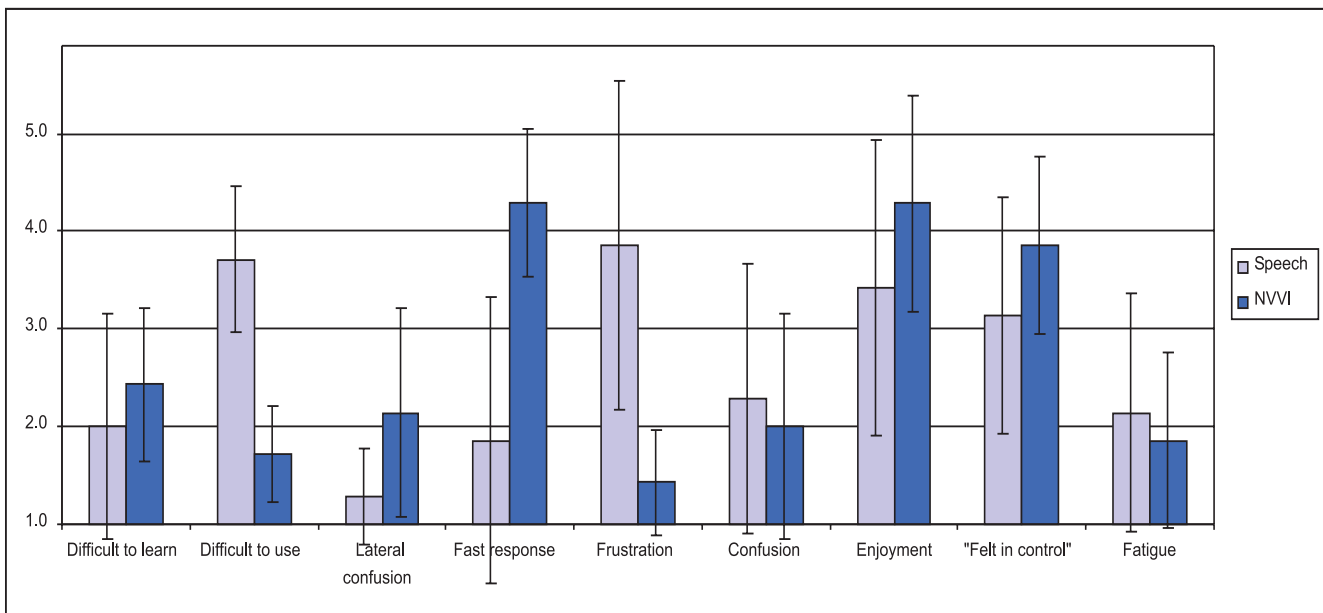
The two-tailed *t*-test ( $p = .05$ ) proved significant differences within the following pair of values: „Difficult to use“ (the speech control was considered more difficult), „Fast response“ (the non-speech sound input was considered providing faster responses than the speech control) and „Frustration“ (the level of frustration when using the non-speech sound input was lower).

The overall grade assigned by the participants on the 1...10 scale (1 worst, 10 best) and their overall comments are reported in Table 3. The ratings of non-speech sound input were significantly better than those of the speech control.

Overall, the non-speech sound input proved a better solution than the control by speech recognition for the real-time control of a model car.

**Table 3: Overall Evaluation (1—10 scale) and user comments**

User	SR	NS	Comment
U1	2	8	"I like the NS method better than speech."
U2	5	8	"The NS method is easier to use."
U3	6	9	"The speech recognition does not work well."
U4	6	9	"The reaction time of NS is much better."
U5	2	8	"No need for the stop command."
U6	1	9	"NS: fastest reaction -- SR: poor recognition."
U7	7	9	"Worked better, was more convenient, and more interesting."



**Fig. 5: Subjective evaluation results. The error bars show the standard deviation.**

## Acknowledgment

Adam Sporka's research is currently being supported by European Commission Marie Curie Excellence Grant for the ADAMACH project (contract. No. 022593). However, the study presented in this paper has been partly supported by the Ministry of Education, Youth and Sports of the Czech Republic under the research program LC-06008 (Center for Computer Graphics).

## References

1. Al-Hashimi S.: Blowttr: A voice-controlled plotter. In Proceedings of HCI 2006 Engage, The 20th BCS HCI Group conference in co-operation with ACM, vol. 2, London, England, pages 41–44, Sept 2006.
2. Bushnell N., Bristow S.: Breakout. A video game, Atari Inc., 1976.

3. Computational Intelligence Group at CTU Prague: Bluecar. <http://ncg.felk.cvut.cz/projects/bluecar/>. Retrieved October 20, 2007.
4. Hamalainen P., Maki-Patola T., Pulkki V., Airas M.: Musical computer games played by singing. In G. Evangelista and I. Testa, editors, Proceedings of 7th International Conference on Digital Audio Effects, Naples, Italy, pages 367–371, 2004.
5. Harada S., Landay J. A., Bilmes J. A.: Drawing with voice: Combining non-verbal vocalizations with words to draw hands-free. In: Proceedings of the CHI 2007 Workshop Striking a C[h]ord: Vocal Interaction in Assistive Technologies, Games, and More, San Jose, California, pages 9–12, 2007.
6. T. Igarashi, J. F. Hughes: Voice as sound: using nonverbal voice input for interactive control. In Proceedings of the 14th annual ACM symposium on User interface software and technology. ACM Press, New York, 2001, 155-156.
7. Pazhitnov A., Gerasimov V., Pavlovsky D.: Tetris. A video game, 1985.
8. Sporka A., Kurniawan S., Mahmud M., Slavík P.: A Longitudinal Study of Continuous Non-Speech Operated Mouse Pointer. In Proceedings of INTERACT 2007, Rio de Janeiro, Brazil. Lecture Notes in Computer Science (LNCS) vol. 4663, Part II., Springer-Verlag, pp. 489 – 502.
9. Sporka A., Kurniawan S., Mahmud M., Slavík P.: Longitudinal Study of Continuous Non-Speech Operated Mouse Pointer. In Extended Abstracts of CHI 2007, San Jose, California, ACM.
10. Sporka A., Slavík P., Kurniawan S. H. Acoustic Control of Mouse Pointer. Universal Access in the Information Society. 2006, vol. 4, no. 3, pp. 237-245. ISSN 1615-5289.
11. Sporka A. J., Kurniawan S. H., Mahmud M., Slavík P. Non-speech Input and Speech Recognition for Real-time Control of Computer Games. In Proceedings of The Eight International ACM SIGACCESS Conference on Computers and Accessibility. New York: ACM Press, 2006, s. 213-220.
12. Sporka A. J., Kurniawan S. H., Slavík P.: Whistling User Interface (U3I). In 8th ERCIM International Workshop “User Interfaces For All”, LCNS 3196, Vienna, pages 472–478. Springer-Verlag Berlin Heidelberg, June 2004.

**About the authors:**



*Adam J Sporka* is a senior-year PhD candidate at the Czech Technical University (CTU) in Prague where he also received his master’s degree in computer science. As a senior research fellow (Marie Currie programme), he has recently joined the University of Trento. In his research, he focuses on speech user interfaces and non-verbal vocal input for emulation of input devices of personal computing equipment. He wrote or contributed to about 25 papers and articles published in scientific journals and proceedings of various international conferences. He was one of the organizers of a first workshop on non-verbal vocal interaction at the ACM CHI 2007 conference. He is also a freelance consultant in HCI and software development. His clients include Czech Academy of Sciences and Prague Philharmonic Choir.



*Pavel Slavík* is a full professor at the Department of Computer Graphics and Interaction at the CTU in Prague. He received his PhD at the Department of Computer Science and Engineering in (1983), where he also received his Associate Professor (1994), and Professor (2003) titles. His wide range of interests includes computer graphics, scientific visualization, graphic user interfaces, and interfaces for users with special needs. He is an author or co-author of more than 200 scientific papers and articles in journals and conference proceedings. He has an extensive record of participation in various research projects, including INCO projects (WISE/E 1995-96, Virtuos 1998 – 2000), 5<sup>th</sup> and 6<sup>th</sup> Framework projects (ENORASI, 2000 – 2001, MUMMY 2001 – 2005, i2home).