Augmentative Communication Based on Realtime Vocal Cord Vibration Detection

Tiago H. Falk, *Member, IEEE*, Julie Chan, Pierre Duez, Gail Teachman, and Tom Chau, *Senior Member, IEEE*

***Abstract*—A binary switch based on the detection of periodic vocal cord vibrations is proposed for individuals with multiple and severe disabilities. The system offers three major advantages over existing speech-based access technologies, namely, insensitivity to environment noise, increased robustness against user-generated artifacts such as coughs, and reduced exertion during prolonged usage periods. The proposed system makes use of a dual-axis accelerometer placed noninvasively in proximity of the vocal cords by means of a neckband. Periodic vocal cord vibrations are detected using the normalized cross-correlation function computed from anterior–posterior and superior–inferior accelerometry signals. Experiments with a participant with hypotonic cerebral palsy show the proposed system outperforming a popular commercial sound-based system in terms of sensitivity, task time, and user-perceived exertion.**

***Index Terms*—Augmentative and alternative communication, normalized cross-correlation, sound switch, vocal cord vibration.**

I. Introduction

WORLDWIDE, approximately 1.3% of all individuals have complex communications needs and cannot rely on natural speech for communication [1]. Such individuals typically possess no functional movements and have severe dysarthria or phonatory impairments as a result of degenerative neuromotor diseases, cerebral palsy, or brain injuries. As an example, upwards of 85% of all individuals with cerebral palsy have severe dysarthria [2]. Due to limitations in available augmentative and alternative communication (AAC) technologies, the majority of these individuals possess no means of communication [3].

Recent advances in speech processing technologies have allowed individuals with severe motor impairments to interact with a computer using speech recognition and nonverbal vocalizations. Representative technologies include the Whistling User Interface [4], Phonetic Control [5], Vocal Joystick [6], Voice Pen [7], and Voice Draw [8] where users control the mouse pointer by changing the pitch, loudness, and vowel quality of their vocalizations, hums, or whistles. This fine phonatory control is obtained by smoothly changing mouth shape and tongue position, and by voluntary control of vocal cord length [9]; such control, however, is not possible for individuals with severe motor impairments affecting voice and speech production.

While phonatory impairments can be associated with poor respiratory control, laryngeal dysfunction, or oral–facial muscular weakness, studies suggest that voluntary phonation can be developed [2]. Produced vocalizations, however, are often unintelligible, pitch invariant, and of reduced loudness [10], thus preclude the use of existing speech-based technologies. In such cases, augmentative communication is often achieved by means of a scanning virtual keyboard (e.g., WiViK [11]) and a binary sound-based switch (e.g., Words+ [12]) which activates once sounds are detected via a close-talking microphone.

As with other speech-based technologies, environment noise, user-generated artifacts (e.g., coughs) [9], and user fatigue [13] play key roles in characterizing system performance. It is customary for sound-based switches to be equipped with a "sensitivity" dial which allows the user to tradeoff noise robustness for low-volume vocalization detection. In noisy environments, loud vocalizations have to be produced for accurate switch activation, hence leading to premature fatigue. In quiet environments and in scenarios involving throat microphones (e.g., [14]), softer vocalizations can be produced. In such instances, however, user-generated artifacts such as coughs, throat clearing, and respiratory noises, as well as the output from the user's speech generating devices, can cause false switch activations [9].

In this paper, we venture away from conventional microphone-based speech AAC technologies and propose a novel binary switch based on periodic vocal cord vibration detection. Speech sounds are produced by forced air from the lungs as it passes between the vocal cords. Voiced speech sounds (e.g., vowels or certain consonants) and humming, for example, cause quasi-periodic vibrations of the vocal cords [15]. Coughs, swallows, and throat clearing, on the other hand, cause aperiodic vibrations. Periodic vocal cord vibrations are detected by means of a normalized cross-correlation function, computed for signals measured from a dual-axis accelerometer placed on the anterior surface of the throat. The proposed solution is insensitive to environment noise, robust to user-generated artifacts, and less strenuous to use, thus overcomes major limitations of existing speech-based technologies and contributes positively to AAC outcomes [16].

The remainder of this paper is organized as follows. Section II describes the developed prototype device. Section III reports

Manuscript received May 07, 2009; revised August 28, 2009; accepted October 19, 2009. First published January 12, 2010; current version published April 21, 2010. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

T. H. Falk and T. Chau are with the Bloorview Research Institute/Bloorview Kids Rehab, Toronto, ON, M4G 1R8 Canada and the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, M5S 3G9 Canada (e-mail: tiago.falk@ieee.org; tom.chau@utoronto.ca).

J. Chan is with the Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, M5S 3G9 Canada.

P. Duez is with the Bloorview Research Institute/Bloorview Kids Rehab, University of Toronto, Toronto, ON, M4G 1R8 Canada.

G. Teachman is with the Bloorview Kids Rehab, Toronto, ON, M4G 1R8 Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2009.2039593

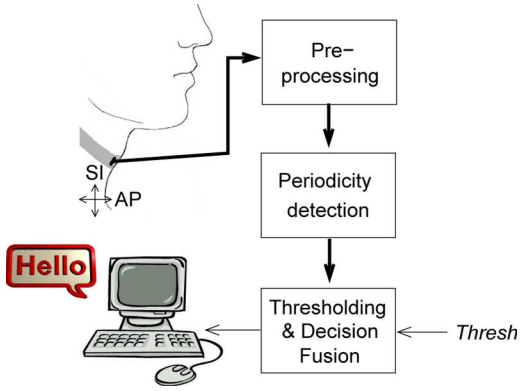


Fig. 1. Block diagram of the proposed system.

comparative results between the device and a commercial sound switch. Section IV presents the conclusions.

II. System Description

Fig. 1 depicts a block diagram of the proposed system. A dual-axis accelerometer is placed on the neck—in proximity of the vocal folds—with the assistance of a neckband. The axes of acceleration are aligned to the anterior–posterior (AP) and superior–inferior (SI) directions, as illustrated by the figure. Both signals are preprocessed and then analyzed by a periodicity detection algorithm. Thresholding is then applied to remove user-generated vibration artifacts caused by coughs or throat clearing. Lastly, decision fusion is performed to account for both AP and SI signal decisions; detected vocalizations can then be used to control a human–computer interface or an AAC device. Individual processing blocks are described in more detail in the subsections to follow.

A. Preprocessing

Both AP and SI signals are sampled at a rate of 1000 Hz (with an anti-aliasing filter in place) and are high pass filtered by a fifth-order Butterworth filter with cutoff frequency at 50 Hz. This choice of frequencies is used in order to place emphasis on the 50–500 Hz range of typical vocal fold vibration frequencies [17]. High pass filtering is also employed in order to remove low-frequency vibrations caused by swallows or (in)voluntary head movements. Representative AP and SI signals are depicted in Fig. 2(a) for vocalization and cough instances.

B. Periodicity Detection

In this paper, the normalized cross-correlation (NCC) function is used to detect periodicity in AP and SI accelerometry signals. More complex statistical based measures (e.g., [18]) may be explored for enhanced performance; such investigation, however, is left for future study. Let L denote frame size duration and $L_w < L$ the correlation window size duration. NCC is defined as

$$\mathrm{NCC}(k) = \frac{1}{\sqrt{e_0 e_k}} \sum_{n=0}^{L_w-1} s(n)s(n+k) \qquad (1)$$

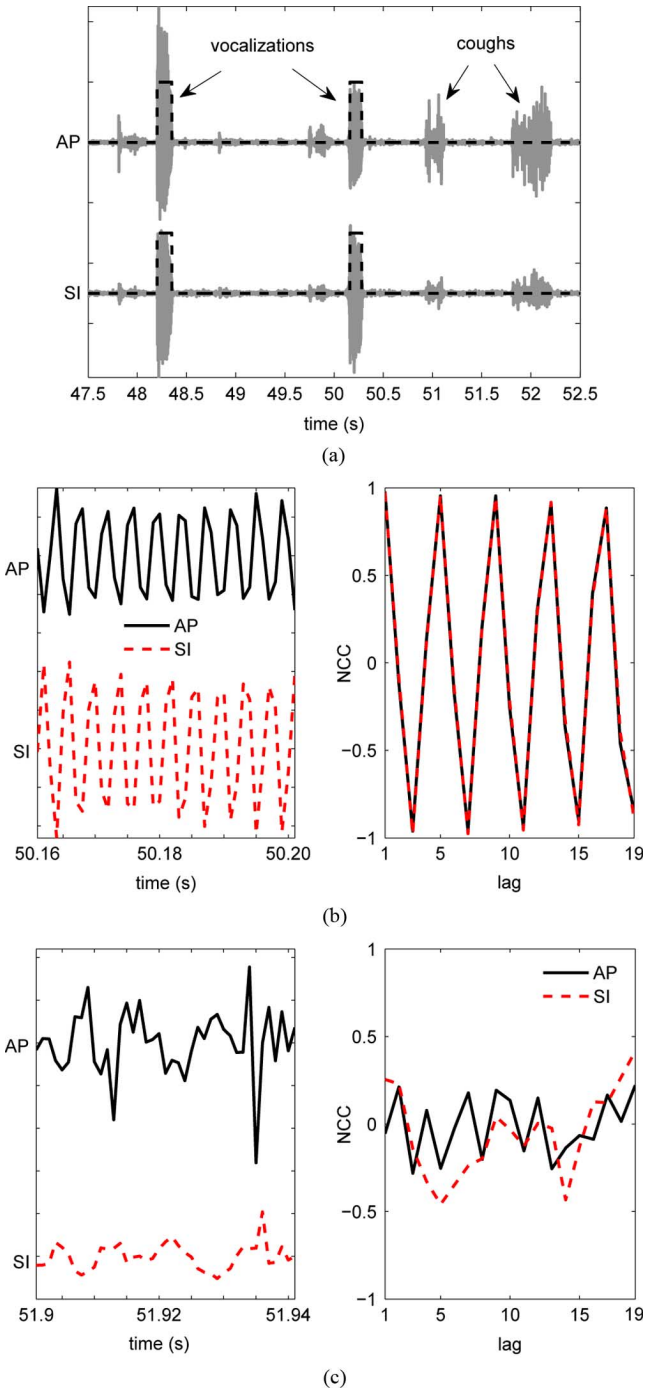


Fig. 2. Plots of (a) AP and SI signals for vocalization and cough instances along with detected periodic events (dashed line), and zoomed-in plots of 40 ms frames (left) and corresponding NCC functions (right) for (b) vocalizations and (c) coughs. For (b), the AP (solid line) and SI (dashed line) curves are superimposed.

where $k = 0, 1, \ldots k_{\max}$, $k_{\max} < L - L_w$ is the maximum lag, $s(n)$ is the zero-mean framed signal segment (AP or SI), and e_k is the energy of the windowed signal and given by

$$e_k = \sum_{n=k}^{k+L_w-1} s^2(n). \qquad (2)$$

The NCC function can assume values between $[-1, 1]$ with values close to unity indicating periodicity.

Motivated by conventional speech processing algorithms (e.g., [19]), signals are analyzed using 40 ms frames ($L = 40$ at a sampling rate of 1000 Hz) and a correlation window size of 20 ms ($L_w = 20$). Plots in Fig. 2(b) and (c) depict 40 ms segments of AP and SI signals (left subplots) and the respective NCC functions (right subplot) for instances of vocalizations and coughs, respectively. As can be seen, vocal fold vibrations caused during vocalization are quasi-periodic, resulting in a periodic NCC function with peak values close to unity. In turn, vibrations caused by unwanted artifacts such as coughs are shown to be aperiodic and to attain low NCC values. It is important to emphasize that other user-generated artifacts, such as movement of the body, may exhibit quasi-periodic behavior; such artifacts, however, have frequencies below 50 Hz and are filtered out during signal preprocessing.

C. Thresholding and Decision Fusion

As observed in Fig. 2(b), periodic signal segments are characterized by a periodic NCC function with cross-correlation values that approach unity near the peaks. Our preliminary experiments have also suggested that coughs can occasionally cause quasi-periodic behavior in AP and SI accelerometry signals. Hence, in order to avoid detection errors, thresholding and decision fusion is applied. Let $\mathbf{p}_{\text{class}} = \{p_{(\text{class},1)}, p_{(\text{class},2)}\}$ denote the first two peaks of the computed NCC function starting from 0 lag [i.e., the two leftmost maxima in the right panels of Fig. 2(b) and (c)]; subscript *class* indicates either AP or SI channels. A vocalization is detected if the following rule holds:

$$\text{Vocalization} = \begin{cases} \text{TRUE}: & \text{if } \{p_{\text{AP},1} > \text{Thresh and } p_{\text{AP},2} > \text{Thresh}\} \\ & \text{or } \{p_{\text{SI},1} > \text{Thresh and } p_{\text{SI},2} > \text{Thresh}\} \\ \text{FALSE}: & \text{otherwise} \end{cases} \quad (3)$$

where Thresh is a user defined threshold that regulates how much "aperiodicity" is accepted by the device. Empirically, we have found $\text{Thresh} = 0.8$ to be an optimal value for distinguishing vocalizations from other sources of vibration (see Section III-C). However, in order to account for individual differences in vocal fold function (e.g., almost "whispered" vocalizations that may be produced when the individual is fatigued), an adjustable threshold dial is implemented in the final hardware solution.

D. Hardware Implementation

A realtime hardware implementation of the proposed system was developed using a PIC microcontroller (PIC24FJ64GA002). A 5 kΩ potentiometer is used to allow for user control of the threshold which was preset to lie between $[0.5, 1]$. The normalized cross-correlation function was programmed into the microcontroller; vocalization detection is performed based on (3). The proposed solution allows for two possible outputs: either a high voltage for usage with conventional interfaces or an F11 keystroke for use with a virtual keyboard such as WiViK [11]. The implemented system along with the accelerometer and custom-made neckband are depicted in Fig. 3.

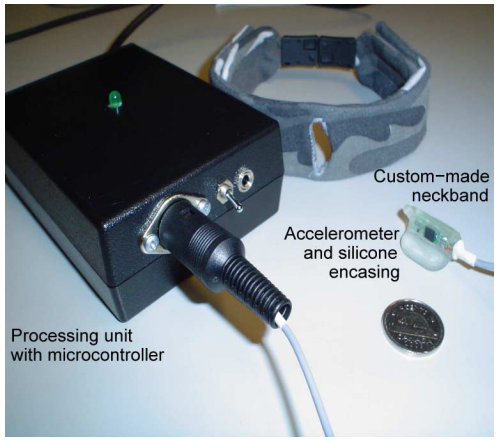


Fig. 3. Implemented switch based on vocal cord vibration detection. System was developed using a PIC microcontroller; accelerometer had a custom-built silicone encasing that was attached to the neckband.

III. Results

In this section, we compare the accuracy of the developed prototype device to that of a commercial off-the-shelf sound switch; results are reported in terms of detection accuracy, time to perform predetermined tasks, and user-perceived fatigue.

A. Participants

Two able-bodied adults (one male and one female) were used to test the sensitivity of the proposed system to varying pitch ranges and to adjust the threshold level. A state-of-the-art pitch tracking algorithm [19] was used to measure average perparticipant pitch values. The male adult had an average pitch of 90 Hz whereas the female had an average pitch of 170 Hz. Two individuals from the target population—one child and one adolescent—also participated in the study and had average pitch values of 280 Hz and 210 Hz, respectively. Both individuals were diagnosed with hypotonic cerebral palsy and have severe phonatory impairments.

B. Performance Metrics

Sensitivity and specificity are used as performance metrics and are given by

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (5)$$

where TP and TN refer to true positives and true negatives, respectively; FP and FN refer to false positives (e.g., detected coughs) and false negatives (e.g., undetected volitional vocalizations). Sensitivity relates to the percentage of correctly identified vocalizations, whereas specificity relates to the percentage of correctly rejected user-generated artifacts. The average time taken to complete the experiment was also recorded and used

for algorithmic comparisons. Metrics relating to user fatigue are described in Section III-D.

C. Threshold Adjustment

The protocol used to adjust the threshold level required able-bodied adults to produce three consecutive vocalizations of the vowel /a/ at three decreasing loudness levels (loud, medium, soft), followed by three swallows and three coughs. Participants in the target population were asked to vocalize as often as possible during a 2.5-min recording session and were also asked to cough, clear their throat, and exhibit respiratory effort several times throughout the experiment. During this pilot experiment, the child vocalized 31 times (with six coughs and throat clearings) and the adolescent vocalized 51 times and produced 19 coughs/throat clearings. For parameter adjustment, the threshold was varied from 0.5–1.0 in increments of 0.05; it was observed that improved performance was obtained with $\mathrm{Thresh} = 0.8$. Using this threshold, 100%, 97%, and 98% accuracy was obtained for able-bodied participants and for the child and adolescent participants in the target population, respectively. Moreover, all coughs and throat clearings were correctly rejected.

D. Comparative Analysis

To compare the performance of the proposed system with that of Words+ [12], an experiment similar to the one described in [14] was conducted with the adolescent participant. The experiment consisted of eight sessions (four in the morning and four in the afternoon) and the participant was asked to copy a pangram sentence containing all letters of the alphabet using the WiViK virtual scanning keyboard (scanning rate set at 1.5 s) [11] with both systems; system usage was altered between days. Due to the nature of the task, the "gold standard" is the known (expected) sequence of timed activations required to write the sentences. Sessions were conducted in a quiet room to avoid any sound switch biases that may have resulted from excessive false activations due to environmental noise. The Words+ sound switch sensitivity dial was set to a maximum value to allow soft vocalizations to be detected. The participant had extensive prior experience with Words+ and WiViK.

Performance results are reported in Table I for both systems. As can be seen, the proposed system attains improved sensitivity relative to Words+, in particular during the afternoon where fatigue is known to hamper the production of loud vocalizations with this particular participant; an improvement of approximately 36% is attained. In terms of specificity, both systems achieved similar performance with Words+ obtaining somewhat lower performance in the afternoon due to a false activation resultant from heavy breathing. In the afternoon, a significant reduction ($p = 0.0038$ using a t-test) in average time for task completion was observed; the task was completed with the proposed system twice as fast relative to Words+.

Three additional metrics are used to quantify user fatigue: total number of rest periods taken during the eight sessions, the average duration per rest period, and the change in perceived exertion before and after the completion of the task. In order to measure user-perceived exertion, a modified five-point Borg scale was used. The participant was asked to rate how much effort was required to complete the task and how tired he felt using a five-point linear scale: [1–Nothing at all, not tired; 2–A little, not tired; 3–Moderate, a little tired; 4–A lot, tired; 5–Too much, very tired]. During each session, the participant was asked to rest between tasks such that pre-task exertion levels remained the same for the two switches. The perceived exertion metric reported in Table I refers to the average difference between post- and pre-task perceived exertion levels; higher values indicate higher levels of fatigue. As can be seen, the participant did not require rest breaks to complete the task with the proposed system and reported substantially lower exertion levels relative to Words+.

TABLE I
PERFORMANCE COMPARISON BETWEEN THE PROPOSED SYSTEM AND THE Words+ SOUND SWITCH FOR AN INDIVIDUAL WITH CEREBRAL PALSY. TEMPORAL METRICS ARE REPORTED AS MEAN ± STANDARD DEVIATION

Performance metric	Words+		Proposed	
	Morning	Afternoon	Morning	Afternoon
Sensitivity (%)	82.4	63.5	85.9	86.4
Specificity (%)	100.0	99.8	100.0	100.0
Average task time (s)	371.6 ± 90.9	783.3 ± 91.4	324.7 ± 47.0	379.6 ± 70.8
Total rest periods	1	8	0	0
Duration/rest period (s)	16.5	18.9 ± 5.8	0	0
Perceived exertion	1.75	2.00	0.50	0.50

Additional testaments to the advantages obtained with the proposed solution include the multiple "chats" carried out with the participant after each session. The participant asked about the authors' preferences of movies, rock bands, and foods. In one instance, he told his educational assistant that a chocolate allergy was the reason for him being sick that day. He also requested a new functionality for WiViK that would enable him to play online video games that require multiple buttons to be pressed at once. According to his educational assistant, this is the most she has seen him interact with his computer and with others. The use of soft vocalizations and hums has also been reported to be less disruptive to his peers at school.

IV. CONCLUSION

A novel binary switch based on periodic vocal cord vibration detection is proposed. The system is shown to overcome three major shortcomings of existing speech-based access technologies—robustness to environment noise, user-generated artifacts, and user fatigue. Additionally, hums can be used for switch activation. Hums are less strenuous on the vocal folds relative to voiced vocalizations, hence play an important clinical factor during prolonged switch usage. The system was tested on four participants (two able-bodied and two clients) and showed reliable performance over a wide range of pitch values. With a participant with hypotonic cerebral palsy, the proposed system outperformed a commercial sound-based solution in terms of sensitivity (36% improvement), average task time (53% reduction), and perceived exertion (87.5% reduction). The single case study results reported here are encouraging and warrant further experiments to determine the usability of the device with other user populations. Further improvements are being investigated to allow for quaternary switch outputs based on discriminating

different pitch frequencies (high or low) and vocalization durations (short or long).

ACKNOWLEDGMENT

The authors wish to acknowledge Mr. K. L. Tam for his assistance with the hardware implementation.

REFERENCES

[1] D. Beukelman and P. Mirenda, *Augmentative and Alternative Communication: Supporting Children and Adults With Complex Communication Needs*, 3rd ed. Baltimore, MD: Paul H. Brookes, 2005.

[2] K. Yorkston, D. Beukelman, E. Strand, and K. Bell, *Management of Motor Speech Disorders in Children and Adults*. Austin, TX: Pro-Ed, 1999.

[3] S. Blain, T. Chau, and A. Mihailidis, "Peripheral autonomic signals as access pathways for individuals with severe disabilities: A literature appraisal," *Open Rehabil. J.*, vol. 1, no. 1, pp. 27–37, 2008.

[4] A. Sporka, S. Kurniawan, and P. Slavik, *Lecture Notes in Computer Science: User-Centered Interaction Paradigms for Universal Access in the Information Society Ch. Whistling User Interface* (U^3I). New York: Springer, 2004, pp. 472–478.

[5] R. Hainisch and M. Platz, "Phonetic control: A new approach for continuous, non-invasive device control using the vocal tract," in *IEEE Int. Conf. Rehabil. Robot.*, 2007, pp. 688–692.

[6] S. Harada, J. A. Landay, J. Malkin, X. Li, and J. A. Bilmes, "The vocal joystick: Evaluation of voice-based cursor control techniques for assistive technology," *Disability Rehabil.: Assistive Technol.*, vol. 3, pp. 22–34, Jan. 2008.

[7] S. Harada, T. S. Saponas, and J. A. Landay, "Voicepen: Augmenting pen input with simultaneous non-linguisitic vocalization," in *Proc. Int. Conf. Multimodal Interfaces*, 2007, pp. 178–185.

[8] S. Harada, J. Wobbrock, and J. Landay, "Voicedraw: A hands-free voice-driven drawing application for people with motor impairments," in *Proc. Int. Conf. Comput. Accessibil.*, 2007, pp. 27–34.

[9] S. Harada, J. Wobbrock, J. Landay, J. Malkin, and J. Bilmes, "Longitudinal study of people learning to use continuous voice-based cursor control," in *Proc. Conf. Human Factors Comput. Syst.*, Boston, MA, 2009, pp. 347–356.

[10] M. Langley and L. Lombardino, "Neurodevelopmental strategies for managing communication disorders in children with severe motor dysfunction," *Pro-Ed Pub.*, 1991.

[11] WiViK. Bloorview Kids Rehab, Toronto, ON, Canada [Online]. Available: http://www.wivik.com

[12] WordsPlus, Infrared/Sound/Touch (IST) Switch [Online]. Available: http://www.words-plus.com

[13] A. Chang and M. Karnell, "Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue," *J. Voice*, vol. 18, no. 4, pp. 454–466, 2004.

[14] G. L. Lancioni *et al.*, "A voice-detecting sensor and a scanning keyboard emulator to support word writing by two boys with extensive motor disabilities," *Res. Developmental Disabilities*, vol. 30, pp. 203–209, 2009.

[15] I. Titze, *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall, 1994.

[16] S. Lund and J. Light, "Long-term outcomes for individuals who use augmentative and alternative communication: Part III—Contributing factors," *Augmentative Alternative Commun.*, vol. 23, no. 4, pp. 323–335, 2007.

[17] M. Hasan, S. Hussaina, H. Setua, and M. Nazrula, "Signal reshaping using dominant harmonic for pitch estimation of noisy speech," *Signal Process.*, vol. 86, no. 5, pp. 1010–1018, 2006.

[18] S. Wichert, K. Fokianos, and K. Strimmer, "Identifying periodically expressed transcripts in microarray time series data," *Bioinformatics*, vol. 20, no. 1, pp. 5–20, 2004.

[19] D. Talkin, "Speech coding and synthesis," in *Ch. A Robust Algorithm for Pitch Tracking (RAPT)*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 495–518.

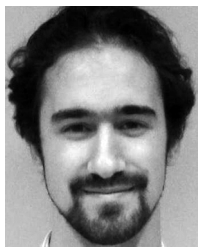
Tiago H. Falk (S'00–M'09) received the B.Sc. degree from the Federal University of Pernambuco, Brazil, in 2002, and the M.Sc. (Eng.) and Ph.D. degrees from Queen's University, Kingston, ON, Canada, in 2005 and 2008, respectively, all in electrical engineering.

He is currently a Postdoctoral Fellow at Bloorview Kids Rehab, affiliated with the University of Toronto, Toronto, ON, Canada. His research interests include biomedical signal processing, rehabilitation engineering, and multimedia quality measurement.

Dr. Falk is recipient of the IEEE Kingston Section Ph.D. Research Excellence Award (2008), the Best Student Paper Awards at ICASSP (2005) and IWAENC (2008), and the Newton Maia Young Scientist Award (2001).

Julie Chan received the B.A.Sc. degree in systems design engineering with an option in biomechanics from the University of Waterloo, Waterloo, ON, Canada, in 2007, and completed a M.H.Sc. in clinical engineering from the University of Toronto, Toronto, ON, Canada, in 2009.

She joined the Paediatric Rehabilitation Intelligent Systems Multidisciplinary Lab at Bloorview Kids Rehab, University of Toronto, Toronto, ON, Canada, for a clinical internship.



Pierre Duez received the B.A.Sc. degree in engineering science and the M.A.Sc. degree in industrial engineering from the University of Toronto, Toronto, ON, Canada, in 2000 and 2003, respectively.

He is a software developer in the PRISM lab at Bloorview Kids rehab, where he supports the implementation of researcher-derived algorithms for access and assistive technologies.

Gail Teachman received the M.S. degree in rehabilitation science from the School of Rehabilitation Science, University of Toronto, Toronto, ON, Canada.

She is a Clinical Associate with the Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada. Her research activities focus on classroom writing, gaining children's perspectives and the measurement of outcomes related to assistive technology. She works at Bloorview Kids Rehab in Toronto in the field of augmentative and alternative communication.

Tom Chau (SM'XX) received the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada.

He is a Senior Scientist at the Bloorview Research Institute and an Associate Professor in the Institute of Biomaterials and Biomedical Engineering at the University of Toronto, Toronto, ON, Canada. Since 2004, he has held a Canada Research Chair in Pediatric Rehabilitation Engineering. He is graduate coordinator of the Clinical Engineering Program and Leader of the NSERC CREATE: Academic Rehabilitation Engineering program at the University of Toronto. His recent research has focused on novel access pathways for children and youth with severe physical impairments. Other research includes the noninvasive monitoring of swallowing and the fractal dynamics of quasi-periodic motor activities.