7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

Video game speech rehabilitation for velopharyngeal dysfunction: Feasibility and pilot testing

Meredith J. Cler, Graham E. Voysey, and Cara E. Stepp

Abstract— **Poor control over the velopharyngeal (VP) port (connection between the oral and nasal cavities) leads to unintelligible speech; this VP dysfunction (VPD) can be due to structural abnormalities, poor motor control, or lack of appropriate feedback (hearing impairment). VP control is not aided by visual feedback since the relevant anatomy is not visible to the speaker or the listener. Here we present initial data from a novel, game-based speech rehabilitation platform designed for children with VPD, in which online feedback of speech nasalization is provided based on measurements of nasal skin vibration and speech acoustics. Twelve pediatric participants (three with VPD) completed one session with the video game and were all able to easily use the game. Over 90% of the participants reported that the game was at least "kind of fun" and that the equipment at least "kind of comfortable". Over 90% of participants and 100% of their parents said they could use the game at home. Results are promising for further development and long-term testing in individuals with VPD.**

I. Introduction

The velopharyngeal (VP) port separates the oral and the nasal cavity and is open during production of the nasal sounds in English: /m/, /n/, and /ŋ/ (e.g., "dim", "din", and "ding"). Vowels produced near nasal sounds are also produced with an open VP port, but the VP port must otherwise be closed in intelligible speech production. A closed VP port during speech that should be nasal results in *hyponasal* speech (e.g. "mom" can sound like "bob", as when a typical speaker has an upper respiratory infection); an open VP port during speech that should not be nasal results in *hypernasal* speech. Poor control over the VP port leads to unintelligible speech. Individuals can present with VP dysfunction (VPD) due to structural abnormalities, poor motor control, or a lack of appropriate feedback (e.g., hearing impairment) [1, 2].

VP control is not aided by typical visual feedback since the relevant anatomy is not visible by the speaker or the listener. Here we present initial data from a novel, game-based speech rehabilitation platform designed for children with VPD, in which online feedback of speech nasalization is provided based on measurements of skin vibration and speech acoustics from a custom miniaturized sensor.

Research supported by the Noonan Foundation and grants DC012651 and DC04663 from the National Institute on Deafness and Other Communication Disorders.

M.J. Cler is with the Graduate Program for Neuroscience-Computational Neuroscience, Boston University, Boston, MA 02215 USA (e-mail: mcler@bu.edu).

G.E. Voysey is with the Graduate Program for Biomedical Engineering, Boston University, Boston MA 02215 USA (e-mail: gvoysey@bu.edu)

C.E. Stepp is with the Departments of Speech, Language, and Hearing Sciences and Biomedical Engineering, Boston University, Boston, MA 02215 USA (phone: 617-353-7487; fax: 617-353-5074; e-mail: cstepp@bu.edu).

II. Methods

A. Participants

Twelve pediatric participants were tested during a single session. Nine of the pediatric participants reported no history of speech, language, or hearing disorders (four female; 4-14 years old, mean: 8.9 years). Two participants were children with hearing impairments and associated VPD (8 year old male; 12 year old female) and one participant was a child with normal hearing and VPD (6 year old male). All were native speakers of American English. Parents completed written consent in compliance with the Boston University Institutional Review Board. Participants aged 7-14 completed verbal or written assent as appropriate. Parents were compensated at $10 per hour and children were compensated with a small toy.

B. Signals

A custom miniaturized sensor was used to measure nasal skin vibration and oral and nasal speech acoustics; these signals were then used to estimate speech nasalization [3]. Combined nasal and oral speech signals were recorded with a WH20 XLR microphone (Shure240 Incorporated, Niles, IL). Nasal skin vibration was recorded with a BU-21771 accelerometer (Knowles Electronics, Itasca, IL) wired to utilize an XLR connection. Both signals were pre-amplified with a RME QuadMic II 4-Channel Microphone Preamp (Audio AG, Haimhausen, Germany) and were digitized via a Komplete Audio 6 external sound card with XLR inputs (Native Instruments, Los Angeles, CA). Nasal accelerometer signals and combined oral and nasal acoustic signals were digitized at a sampling rate of 44.1 kHz using custom C# code. The signals were stored as 256 two-channel, 16-bit WAV files for post hoc analysis. An example set of signals is shown in Fig. 1.

C. Software Architecture

The game was developed in C# using the .NET 4.5.1 framework (Microsoft, Redmond WA). The application was divided into a configuration window (see Fig. 2A) and a full-screen gameplay window (see Fig. 2B-D). A flow diagram is shown in Fig. 3.

In the configuration window, users were required to add or select a participant. Required participant information

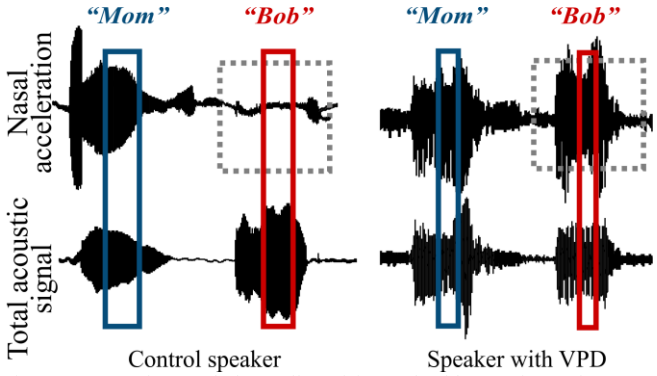

Figure 1. Example speech data collected from a healthy speaker and a speaker with VPD. Signal from the nasal accelerometer shown on top, with combined oral and nasal acoustic output shown on bottom. Middle 30% of the utterance highlighted in blue (nasal word) and red (non-nasal word). Note a clear difference in the nasal accelerometer signal during the non-nasal sound "Bob" (indicated with gray dotted box). The speaker with VPD shows nasal acceleration during this non-nasal word.

included age, gender, and a unique subject ID. Optionally, a name and participant notes could be added. Users were also required to test the microphone and nasal accelerometer in the configuration window before the game could be played. Finally, users selected whether gameplay should show no feedback (i.e. evaluate user's typical nasalization for both nasal and non-nasal words) or give the user feedback based on whether the online estimation of speech nasalization was above or below the nasalization thresholds set in the configuration window (see Fig. 2A).

Once the configuration requirements were met, the user was allowed to select "Play" and begin gameplay in a full-screen window (see Fig 2B-D). A series of consonant-

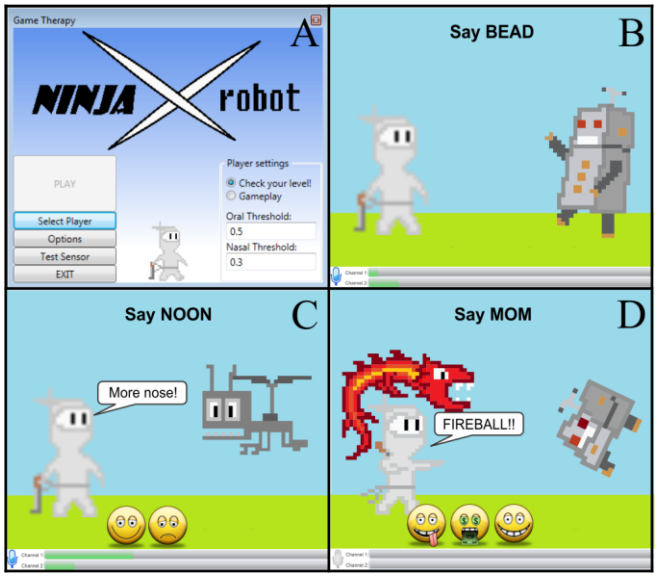

Figure 2. Example screenshots from the game. A: The configuration window, in which users select the username, test the instrumentation, and select whether the game should estimate the user's typical speech nasalization ("Check your level") or provide feedback ("Gameplay") based on the thresholds entered. B: Example in which the user is prompted to repeat the non-nasal word "bead" during the no feedback stage. C: Example in which the user given feedback during a nasal word "noon" during the feedback stage. D: Example of ninja action after three successful repetitions in the feedback stage.

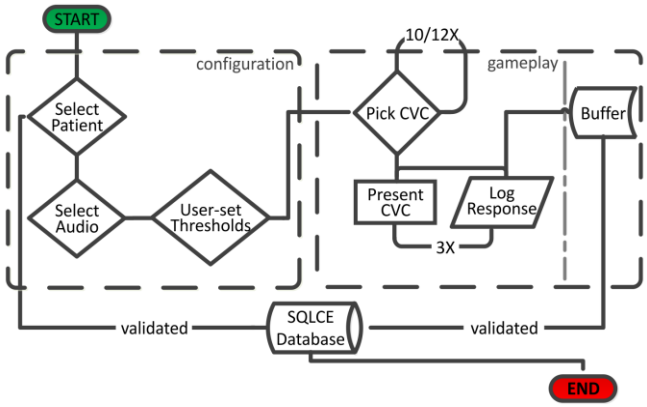

Figure 3. A flow diagram of the configuration window, gameplay, and database. CVC = consonant-vowel-consonant word.

vowel-consonant words were presented to the participants; users were prompted to repeat each three times. After each prompt, 2-s clips of signals from the microphone and accelerometer were recorded as dual-channel uncompressed WAVs, processed online (see *Online signal processing*), and buffered along with relevant metadata for later database storage. Metadata included a timestamp, threshold values, and performance relative to the current threshold.

All gameplay data, including recordings of participant responses, were logged and securely stored in a local SQLCE 4.0 (Microsoft, Redmond WA) standalone database (see Fig. 4). The database was protected with the AES128

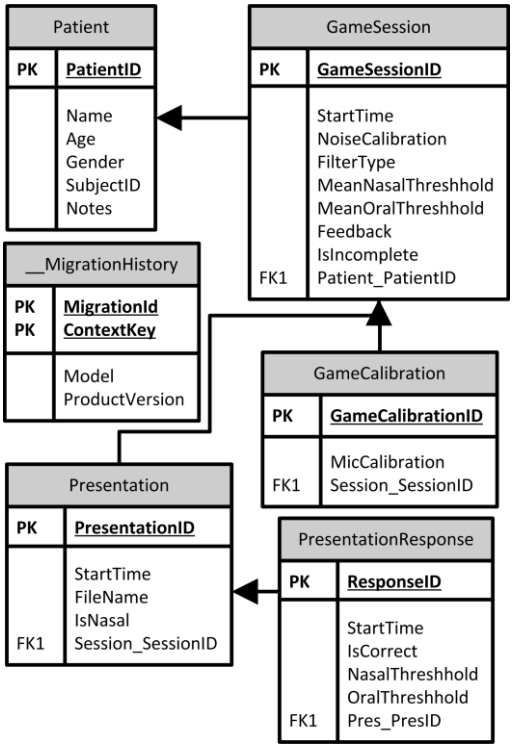

Figure 4. Entity-Relationship Diagram for the game database. Table name is shown in header, and primary keys are denoted with bold, underline, and "PK". Foreign keys are denoted by "FK1". Relationships between entities are indicated with linked arrows; all relationships were specified as One-to-Many. The __MigrationHistory table allows for refinement of the database schema in future versions of the game while preserving automatic backwards compatibility with existing database copies.

encryption and SHA256 hash standards in order to control access to recordings of participants' voices. To guarantee uniqueness of each record stored across multiple parallel database instances, Globally Unique Identifiers (GUIDs) were used for key values.

At the end of each gameplay session, the buffered data were written into the database (see Fig. 4). In the event that the session ended before all words were presented, either by the request of the participant or accidentally, the buffered data were still written into the database, and the session type field was updated to indicate that the session was incomplete.

D. Online signal processing

After the user was prompted to repeat a word, two seconds of acoustic signals were obtained from both the accelerometer and the microphone. A production was extracted from the entire two second recording, determined as the longest period in which the combined oral and nasal signal was above the noise floor. The center 30% of this production was extracted and used to calculate the speech nasalization of the vowel production.

Normalized Horii Oral-Nasal Coupling (HONC) scores were used to quantify nasalization [4]. Nasal accelerometer signals were bandpass filtered between 400 and 1000 Hz, and combined nasal and oral acoustic output was bandpass filtered between 25 and 420 Hz [3, 5]. For each extracted vowel token ($_x$), the normalized HONC score (nH_x) was computed as the ratio of the root mean square (RMS) of the filtered accelerometer signal (A_x) over the RMS of the filtered microphone signal (M_x). This was then normalized by the HONC score calculated from a sustained "mmm" sound (/m/) during a calibration period (i.e. $A_{/m/} / M_{/m/}$).

$$nH_x = (A_x / M_x)/H_{/m/} \quad (1)$$

E. Gameplay

Users were told that they would play the game as a ninja fighting evil robots. Gameplay was divided into two stages: no feedback and feedback. During the first stage, the user was given no feedback, and all productions were used to estimate the user's typical speech nasalization. The second stage began with instructions about how to make nasal and non-nasal speech appropriately nasal and gave the users feedback based on the online calculation of the speech nasalization of each production.

1) No feedback stage

This first stage began with a brief calibration period wherein users were asked to remain completely silent (to measure the noise floor) and then to say "mmm" twice (used to calculate $H_{/m/}$ in Eq 1). Regular gameplay then began and continued for twelve trials. The participant was represented by a ninja avatar on the left of the screen (see Fig. 2B). For each trial, a robot appeared on the right of the screen, and participants were presented with visual cues (e.g. the text "Say BEAD" appeared, and the microphone on the lower left of the screen turned blue; see Fig 2B) and auditory cues (a voice saying "Say bead" and a beep at the correct time) to say one of the stimuli words. They repeated the given stimulus (see Table I) three times in their typical voice as the robot approached the ninja. After the three repetitions, the ninja performed one action pseudorandomly: jumped, ducked, used a sword, or threw a fireball at the robot. After twelve trials, users were told that they did a great job and deserve a treat. The ninja avatar was then shown holding an ice cream cone.

TABLE I. NASAL AND NON-NASAL STIMULI

Stimuli	
Nasal	***Non-Nasal***
Man	Bag
Mean	Bead
Mine	Guide
Nun	Bug
Mom	Dog
Noon	Dude

At the end of this stage, all of the nasal responses and non-nasal responses were sorted separately, and two nasalization thresholds were set such that the user would have produced speech above the nasal or below the non-nasal thresholds 70% of the time using their typical speech [6]. These thresholds were automatically populated in the configuration window before the next stage (see Fig. 2A).

2) Feedback stage

The user was told that the robots had learned not to be afraid of the user's normal voice. Instead, the user needed to fight the flying robots (paired with nasal words) by making their speech all come out of their nose (i.e. increase their speech nasalization and thus the normalized HONC score). They must fight walking robots (paired with non-nasal words) by making their speech all come out of their mouth (i.e. decrease their speech nasalization).

The second stage of gameplay was identical to the first. A robot appeared and the user was prompted with both visual and auditory cues to repeat the given stimulus three times. After each of the user's repetitions, an online estimation of the normalized HONC score (nH_x) of that utterance was calculated and compared to the appropriate nasal or non-nasal threshold set after the first stage. A repetition was deemed successful if a nasal word had a higher nH_x than the nasal threshold or if a non-nasal word had a lower nH_x than the non-nasal threshold. After each repetition, feedback was given to the users: successful repetitions were followed by a happy face, whereas unsuccessful repetitions were followed by a frowning face. After the three repetitions, the ninja reacted to the robot. If all three repetitions were unsuccessful, the ninja said "ouch". If one repetition was successful, the ninja jumped or ducked to avoid the robot. If

two repetitions were successful, the ninja used a sword on the robot. If all three repetitions were successful, the ninja threw a fireball at the robot (see Fig. 2D).

F. Feedback from users

After the two game stages, pediatric users and parents completed questionnaires about how fun and comfortable the game was and whether they thought they could use the game at home. Possible responses were represented by a five point Likert scale with text and visual representations ranging from smiling faces to frowning faces. When available, parents also filled out surveys; parent surveys used the same Likert scale, and the questions were reworded to ask how the parent perceived the child's experience (e.g. "How comfortable did you think the game was for your child?").

III. Results

Overall, participants responded well to the game (see Fig. 5). Nearly 92% of the participants thought the game was "very fun" or "kind of fun", with only one healthy child reporting that the game was "sort of boring". The same proportion thought the game was "comfortable" or "sort of comfortable". All three children with VPD reported the game as "very fun". One child with hearing impairment said "Hey, I should get this game." Another hearing-impaired child reported that the game helped her understand the difference between nasal and non-nasal words. Importantly, 100% of the parents surveyed indicated that they thought their child could use the game at home.

IV. Discussion

Future game development will incorporate feedback from these early users. Hearing-impaired users requested clearer visual cues for when words should be repeated (see microphone color in Fig. 2B versus 2D, indicating when to repeat words). This will also assist younger participants, some of whom struggled with the timing of when to say the words even with visual and auditory cues. Future versions will also display both the cue word (e.g. "mom") and an image depicting the word to assist pre-literate users. After this additional development, feasibility testing of the system in children with VPD will allow assessment of the potential short- and long-term gains in speech function possible with this type of intervention.

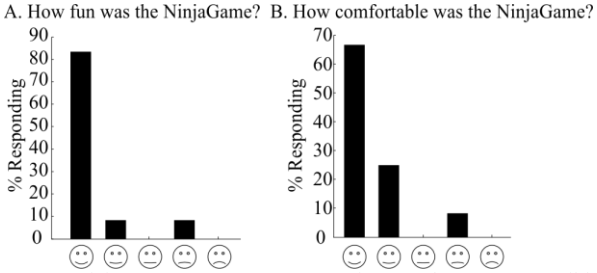

Figure 5. Participant responses to (A) "How fun was the game?" Possible responses: very fun, kind of fun, not fun but not boring, sort of boring, really boring. (B) "How comfortable was the game?" Possible responses: very comfortable, kind of comfortable, not comfortable but not uncomfortable, soft of uncomfortable, and really uncomfortable.

While providing an alternative form of feedback (i.e. nasalization information via visual instead of auditory feedback) is particularly advantageous in individuals with VPD due to hearing disorders, cross-modal feedback has been used in a wide range of speech rehabilitation in individuals without hearing impairments [see e.g. 7]. Our ultimate goal is to facilitate increased appropriate nasalization practice by optimizing motivation, ease of adoption, and perceptual-motor learning. Frequent practice is essential to motor learning [8], and the use of a game format for practicing this sensorimotor skill may increase patient compliance. Video game use in particular may also increase motor-learning effects, as striatal dopamine release during video game play may facilitate brain plasticity following perceptual learning [9, 10].

Most of our participants and all surveyed parents believed that they could use this video game at home, which would thus provide more practice and more targeted practice than would be available outside of speech therapy. Our results provide encouraging evidence that combining VP port biofeedback with a video game environment may encourage increased learning during rehabilitation and improve the intelligibility of speech. The modular structure of the game is intended as a framework for providing online feedback to rehabilitate a variety of motor-learning disorders.

Acknowledgments

The authors would like to thank Dave Anderson, Maia Braden, Andrew Brughera, Will Cunningham, Ran Gong, Jake Hermann and Lenny Varghese for their assistance.

V. References

[1] M. B. Higgins, A. E. Carney, and L. Schulte, "Physiological assessment of speech and voice production of adults with hearing loss," *J Speech Hear Res,* vol. 37, pp. 510-21, Jun 1994.

[2] A. Ysunza and M. C. Vazquez, "Velopharyngeal sphincter physiology in deaf individuals," *Cleft Palate Craniofac J,* vol. 30, pp. 141-3, Mar 1993.

[3] L. A. Varghese, J. O. Mendoza, M. N. Braden, and C. E. Stepp, "Effects of spectral content on Horii Oral-Nasal Coupling scores in children," *J Acoust Soc Am,* vol. 136, p. 1295, Sep 2014.

[4] Y. Horii, "An accelerometric approach to nasality measurement: a preliminary report," *Cleft Palate J,* vol. 17, pp. 254-61, Jul 1980.

[5] E. B. Thorp, B. T. Virnik, and C. E. Stepp, "Comparison of nasal acceleration and nasalance across vowels," *J Speech Lang Hear Res,* vol. 56, pp. 1476-84, Oct 2013.

[6] N. A. Macmillan and C. D. Creelman, *Detection theory: a user's guide*. Cambridge England; New York: Cambridge University Press, 1991.

[7] J. L. Preston, N. Brick, and N. Landi, "Ultrasound biofeedback treatment for persisting childhood apraxia of speech," *American Journal of Speech-Language Pathology,* vol. 22, pp. 627-643, 2013.

[8] J. Robbins, S. G. Butler, S. K. Daniels, R. Diez Gross, S. Langmore, C. L. Lazarus, B. Martin-Harris, D. McCabe, N. Musson, and J. Rosenbek, "Swallowing and dysphagia rehabilitation: translating principles of neural plasticity into clinically oriented evidence," *J Speech Lang Hear Res,* vol. 51, pp. S276-300, Feb 2008.

[9] S. Bao, V. T. Chan, and M. M. Merzenich, "Cortical remodelling induced by activity of ventral tegmental dopamine neurons," *Nature,* vol. 412, pp. 79-83, Jul 5 2001.

[10] M. J. Koepp, R. N. Gunn, A. D. Lawrence, V. J. Cunningham, A. Dagher, T. Jones, D. J. Brooks, C. J. Bench, and P. M. Grasby, "Evidence for striatal dopamine release during a video game," *Nature,* vol. 393, pp. 266-8, May 21 1998.