Snorble: An Interactive Children Companion

Mike Rizkalla¹, Thomas Chan¹, Emilio Granell², Chara Tsoukala², Aitor Carricondo², Carlos Bailon², María Teresa González², Vicent Alabau²

¹Snorble, New York, USA
²Sciling S.L., Alboraya, Spain

{mike,thomas}@snorble.com, {egranell,ctsoukala,acarricondo,cbailon,tgonzalez,valabau}@sciling.com

Abstract

This paper presents an interactive companion called Snorble, created to engage with children and promote the development of healthy habits under the Snorble project.

Snorble is a smart companion capable of having a conversation with children, playing games, and helping them to go to sleep, all made possible thanks to speech recognition.

Index Terms: speech recognition, human-computer interaction, children speech recognition

1. Overview of the Snorble Project

Snorble ¹ is a start-up from New York that develops technology and products based on Artificial Intelligence. Their star product is an interactive smart companion (see Figure 1) that can be customized and aims to help small children develop a proper nighttime routine, healthy habits, social-emotional intelligence and establish a love of learning through a gamified path.

In addition to the smart companion, the Snorble project includes an entire cloud infrastructure created to manage the multimedia content that is played or created on Snorble. Also included is an application [2] for mobile devices (see Figure 2) that facilitates Snorble’s configuration.

2. The Snorble doll

Snorble is an engaging smart companion (see Figure 1). The product is an animated character brought to life which children can hug, interact, and even share stories with. The main goal of Snorble is gamifying children’s routines throughout the day with educational activities and healthy habit building actions like teeth brushing. Snorble aims to make the process of learning habits easier for parents and fun for children, while getting a good night’s sleep in the process.

The main operation of Snorble relies on a state machine, which makes it able to interact with children and adults through voice commands and buttons, and even react to the environment thanks to a variety of sensors and actuators. Based on this concept, we are able to design games to interact with Snorble. An example is the “Color Game”, in which Snorble’s face will change its color, asking the child to guess the name of that color. With a simple dynamic, if the child says the right color name, Snorble will congratulate and ask if the child wants to continue playing, in order to show another color. On the other hand, if the provided answer is wrong or a predefined timeout is reached, Snorble will either ask again for the color name or tell the answer.

To control these dynamics and routines, a mobile phone application is available for parents. Through a variety of settings,
parents are able to select, for example, when the bedtime routine should start, or when it is a good time to play, brush their teeth, etc.

The following sections describe the main features of Snorble: the routines (Section 3), the sensors and actuators (Section 4), and the speech recognition interface (Section 5). Finally, Section 6 discusses the work presented and the possible future work lines.

3. Routines

The Snorble’s behavior is defined by the routines, which contain the information needed by the state machine to operate. Each routine is composed by a set of steps, which are linked through transitions. Steps define how Snorble interacts with the child. For example, depending on the step, playing content in the smart companion’s screen or requiring a specific input to continue.

Each routine has one or more voice commands that trigger its first step, after which the state machine starts to run the next steps when possible. Among the possible inputs that may be required to trigger a subsequent step and enhance the interaction with the children, we can find voice commands, specific sensor values, or timers. All of them can be combined through several operators, allowing for a complex structure of steps and transitions, which results in a richer interaction, close to a real conversation.

For example, the aforementioned “Color Game” is implemented through a routine. In a first step, Snorble asks the child to say the name of a particular color. Then, a transition starts with three possible options: right answer, wrong answer and timeout. Depending on the option triggered, the routine continues through a different branch, either congratulating the child, or asking to try again.

4. Sensors and Actuators

The Snorble smart companion contains several sensors and actuators, such as a radar chip, an accelerometer and a haptic device. The sensor/actuator system is based on a hierarchical structure of abstraction layers. It transforms the raw data obtained through the physical sensors placed in the smart companion into human-readable information. Each of the so-called “sensors” reads the raw data of one or more devices (physical sensors of the smart companion) and combines them to obtain an information feature. These features are the ones intended to be used by Snorble routines when reading sensor information.

With that, sensors are capable to monitor the environment: light, noise, temperature, nearby elements, etc. For example, with the radar chip, Snorble can detect the movement of the child that is using it (crawling, running, etc.); with the accelerometer, it can capture the intensity and direction of the smart companion’s motion (left, right, up, down). These features also provide a bidirectional way of playing and communicating with Snorble. For example, if the device’s battery is low it may not want to play with the child, and it will tell the child to put it on the base for charging.

The main information of the sensors is also sent to the mobile application, allowing for a real-time monitoring of Snorble’s status (battery level, smart companion motion, etc.)

5. Speech Interface

One of the biggest software challenges in the Snorble project is the speech recognition interface [3] since the target users are children from newborn to preschool age (0-5 years old) that have their own way of communication [4], that oftentimes only their parents fully understand what they are trying to say. To offer the best user experience the speech interface uses both continuous speech recognition and speech classification. The main continuous speech system is based on Vosk [5] with models trained using Kaldi [6] and the secondary speech classification system is based on Wav2Vec [7]. Moreover, the speech interface is prepared to recognize the speaker between the family members. Currently, the main limitation is that it is only available for the English language.

6. Discussion and Future Work

The Snorble project is a very ambitious one, not only because of the characteristics of the target users (children from 0 to 5 years old) but also because of the entire infrastructure needed to carry it out, both hardware- and software-wise.

Regarding speech recognition in young children, we continue to investigate how to improve performance, such as by preprocessing the audio signal, testing other acoustic models or network architectures, and other speech classification techniques. In the near future the speech recognition system will be prepared for more languages, starting with French, Spanish and Portuguese to complete the American market, to later ideally cover all 24 official languages of the European Union.

7. Acknowledgements

The authors would like to acknowledge the technical contributions of: Ihab Alkhoury, Carlos Millán, Javier Naranjo, David Camino and many others who have participated in some way during the development of the project.

8. References