A Talk on the Wild Side: The Direct and Indirect Impact of Speech Recognition on Learning Gains.

Holmes, W.[1], Mavrikis, M.[1], Rummel, N.[2],
Grawemeyer, B. [3], Wiedmann, M. [2], & Loibl, K. [2]

[1] London Knowledge Lab, Institute of Education, University College London;
[2] Ruhr-Universität Bochum; [3] London Knowledge Lab, Birkbeck, University of London

Abstract

Research in the learning sciences and mathematics education has suggested that 'thinking aloud' (verbalization) can be important for learning. In a technology-mediated learning environment, speech might also help to promote learning by enabling the system to infer the students' cognitive and affective state so that they can be provided a sequence of tasks and formative feedback, both of which are adapted to their needs. For these and associated reasons, we developed the *iTalk2Learn* platform that includes speech production and speech recognition for children learning about fractions. We investigated the impact of *iTalk2Learn*'s speech functionality in classrooms in the UK and Germany, with our results indicating that a speech-enabled learning environment has the potential to enhance student learning gains and engagement, both directly and indirectly.

Introduction

Research in the learning sciences and in mathematics education highlights the important role that speech plays in learning in general and in mathematics in particular. For example, researchers have shown multiple benefits of 'thinking aloud' (verbalization) for learning (e.g. Mercer, 1995) and have suggested that spoken reflection is a key strategy for stimulating retention for later recall (e.g., Freudenthal, 1981). Other research (e.g. Rajala *et al.*, 2012) has shown that, when students are encouraged to give self-explanations about a target mathematical principle, their learning of that principle is enhanced.

Previous work in technology-mediated learning environments (e.g. LISTEN: Mostow & Aist, 2001; ITSPOKE: Litman & Silliman, 2004; Autotutor: D`Mello *et al.*, 2011) has suggested that speech might help to promote learning in at least three other interrelated ways: (1) speech might provide a natural interface beneficial to learning and might also allow learners who have not yet mastered written language to interact more easily, (2) what students say might be used to infer their cognitive state so that they can be provided an adaptive sequence of tasks and appropriate formative feedback, and (3) how students speak might be used to infer their affective state so they can be provided with affect-aware support.

Drawing upon the summative evaluation of a 3-year EU-funded research project, *iTalk2Learn* (FP7 grant agreement #318051), this paper investigates the direct and indirect impacts on learning of using speech recognition and speech production in an adaptive digital learning platform.

Method

Participants

The participants were all children aged between 8 and 12 years old, recruited from three schools in the UK (N = 117) and six schools in Germany (N = 159). Only children whose parents gave consent were included.

Materials

The study involved an adaptive digital learning platform, *iTalk2Learn*, developed to support children learning fractions. The system's intervention model (Mazziotti *et al.*, 2015), which was developed using a design-based research methodology (Design-Based Research Collective, 2003) and a series of Wizard of Oz studies (Mavrikis *et al.*, 2014), combines a novel exploratory learning environment developed by the project (*Fractions Lab*: Hansen *et al.*, 2015) with pre-existing structured practice environments (*Maths Whizz* in the UK and *Fractions Tutor* in Germany). Formative data was used to train Bayesian networks that determine affect-aware intelligent formative feedback strategies (*c.f.* Grawemeyer *et al.*, 2015).

The system incorporates speech production functionality, with the aim of encouraging students to talk to the system because it talks to them. It also incorporates speech recognition functionality, trained with a corpus of children's speech, with the aim of detecting indications of the student's cognitive and affective states. This is used to provide a sequence of tasks and formative feedback, both of which are adapted to the student's needs.

Procedure

Quasi-experimental studies were undertaken in authentic classrooms (i.e. in the 'wild') in the UK and in Germany in order to investigate the hypothesis: *an adaptive system with speech enhances learning more than an adaptive system without speech*. This paper reports data from two of three experimental conditions. The first (the *speech condition*) used the full *iTalk2Learn* platform including speech functionality. In the second (the *non-speech condition*), the speech functionality was switched off. Students were randomly allocated to conditions.

In each study, participating students completed a pre-test, a 40 minute session engaging with the *iTalk2Learn* system, and a post-test. Isomorphic versions of the pre- and post-tests (online questions presented to the students as being integral to the system and designed to assess procedural and conceptual knowledge of fractions) were assigned randomly (internal consistency at pre-test was α_{UK} = .58, α_{DE} = .41, and at post-test α_{UK} = .53, α_{DE} =.42).

In each condition, the students began by answering fractions tasks in *Fractions Lab*. In the *speech condition*, as they constructed their answer using the available fractions representations and tools, students were encouraged to 'think aloud'. At the same time, they were provided formative feedback adapted by means of a Bayesian network to their individual affective state, which was inferred from their speech and interaction (Grawemeyer *et al.*, 2015): speech recognition was used to detect keywords associated with particular affective states (Grawemeyer *et al.*, 2014), interaction data included whether or not feedback had been followed. A 'student needs assessment' component (Mazziotti *et al.*, 2015) then determined the next task to be given to the student. This was based on whether they were under-, appropriately or over-challenged, which in turn was inferred from an analysis of prosodic cues (e.g. the length of a student's spoken vowels and consonants) (Janning *et al.*, 2015) and the amount of feedback that had been provided.

In contrast, in the *non-speech condition*, formative feedback was based only on the student's task performance (Holmes *et al.*, 2015), while the SNA determined the next task based only on the amount of formative feedback that had been provided.

Results

In the UK study, an ANOVA with time of measurement as the within-subjects factor and condition as the between-subjects factor revealed that learning gains in the *speech condition* (d = .75) were higher than in the *non-speech condition* (d = .44), but this difference was not statistically significant, $F(1,115) = 2.762$, $p = .099$, $\eta_p^2 = .023$. In the German study, a similar analysis showed that learning gains in the *speech condition* (d = .75) were also higher than in the *non-speech condition (d = .69)*, although again this was not statistically significant, $F(1,157) < 1$, $p = .727$, $\eta_p^2 = .001$. Anecdotal evidence from class observations and interviews of a subsample of students (N=12) further suggested that the students were more engaged by the *speech condition*.

Discussion

We investigated the hypothesis that an adaptive digital platform with speech functionality enhances learning more than the same system without speech functionality. In fact, while neither the UK nor German result was statistically significant, the students' learning outcomes did appear to benefit from the speech functionality. Encouraging students to speak during learning, to 'think aloud', and using that speech to help infer indications of the student's cognitive and affective states, in order to determine an appropriate sequence of tasks and appropriate formative feedback, did appear to contribute both to learning gains and to student engagement.

Further research is now needed, both to test the reliability of these results and to tease them apart. In particular, we are interested in investigating both the *direct* impact on learning gains and engagement of speech functionality that is used to encourage a student to verbalize their thoughts, and the *indirect* impact on learning gains and engagement of speech functionality that is used to detect and adapt to a student's cognitive and affective state.

References

D`Mello, S. K., Dowell, N., & Graesser, A. (2011). Does it Really Matter Whether Student`s Contributions Are Spoken Versus Typed in an Intelligent Tutoring System With Natural Language? *Journal of Experimental Psychology: Applied 2011*, *17*(1), 1–17.

Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, *32*(1), 5–8.

Freudenthal, H. (1981). Major problems of mathematics education. *Educational Studies in Mathematics*, 133–150.

Grawemeyer, B., Mavrikis, M., Hansen, A., Mazziotti, C., & Gutierrez-Santos, S. (2014). Employing Speech to Contribute to Modelling and Adapting to Students' Affective States. In *Open Learning and Teaching in Educational Communities* (pp. 568–569). Springer.

Grawemeyer, B., Mavrikis, M., Holmes, W., & Gutiérrez-Santos, S. (2015a). Adapting Feedback Types According to Students' Affective States. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (Vol. 9112, pp. 586–590). Springer.

Hansen, A., Mavrikis, M., Holmes, W., & Geraniou, E. (2015). Designing interactive representations for learning fraction equivalence. In N. Amado & S. Carreira (Eds.), *12th International Conference on Technology in Mathematics Teaching* (pp. 395–402). Faro, Portugal.

Holmes, W., Mavrikis, M., Hansen, A., & Grawemeyer, B. (2015). Purpose and Level of Feedback in an Exploratory Learning Environment for Fractions. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (Vol. 9112, pp. 620–623). Springer.

Janning, R., Schatten, C., & Schmidt-Thieme, L. (2015). How to Aggregate Multimodal Features for Perceived Task Difficulty Recognition in Intelligent Tutoring Systems. Proceedings of the 8th International Conference on EDM.

Litman, D. J., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. In *Demonstration papers at HLT-NAACL 2004* (pp. 5–8). Association for Computational Linguistics.

Mavrikis, M., Grawemeyer, B., Hansen, A., & Gutierrez-Santos, S. (2014). Exploring the Potential of Speech Recognition to Support Problem Solving and Reflection. Wizards Go to School in the Elementary Maths Classroom. In *Proceedings of the 9th EC-TEL 2014*.

Mazziotti, C., Holmes, W., Wiedmann, M., Loibl, K., Rummel, N., Mavrikis, M., … Grawemeyer, B. (2015). Robust Student Knowledge: Adapting to Individual Student Needs as They Explore the Concepts and Practice the Procedures of Fractions. Workshop paper presented at the 17th International Conference on Artificial Intelligence in Education, Madrid.

Mercer, N. (1995). *The Guided Construction of Knowledge: Talk Amongst Teachers and Learners*. Clevedon: Multilingual Matters.

Mostow, J., & Aist, G. (2001). Evaluating tutors that listen: An overview of Project LISTEN. In K. F. & P. Feltovich (Ed.), *Smart Machines in Education* (pp. 169–234). MIT/AAAI Press.

Rajala, A., Hilppö, J., & Lipponen, L. (2012). The Emergence of Inclusive Exploratory Talk in Primary Students' Peer Interaction. *International Journal of Educational Research*, *53*, 55–67.