6-18-2022

A speech-based empathy training system - initial design insights

Florian Weber
Information Systems and Systems Engineering, weber@uni-kassel.de

Sebastian Schween
Information Systems and Systems Engineering, sebastian.schween@student.uni-kassel.de

Thiemo Wambsganss
Swiss Federal Institute of Technology in Lausanne (EPFL), thiemo.wambsganss@epfl.ch

Matthias Soellner
University of Kassel, soellner@uni-kassel.de

A SPEECH-BASED EMPATHY TRAINING SYSTEM - INITIAL DESIGN INSIGHTS

Research in Progress

Florian Weber, University of Kassel, Kassel, Germany, weber@uni-kassel.de

Sebastian Schween, University of Kassel, Kassel, Sebastian.Schween@student.uni-kassel.de

Thiemo Wambsganss, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, thiemo.wambsganss@epfl.ch

Matthias Söllner, University of Kassel, Kassel, Germany, soellner@uni-kassel.de

Abstract

Empathy is an essential component of human communication since it increases our understanding and perception of others. However, studies show that students' empathy skills have declined rapidly in the last decades. Against this background, practitioner reports predict that the importance of empathy will increase as a skill for successful agile teamwork in the future. Therefore, researchers have designed information systems to train empathy abilities of learners in different domains. Nevertheless, research on automated speech-based training is rather scarce. Hence, we aim to investigate how to design a speech-based empathy training system that helps students react emotionally adequately in communication. This research in progress paper presents five initial requirements that guide future research and development of a speech-based empathy training system intended to support students' self-regulated learning. With this, we hope to provide guidance for the design and embedding of speech-based empathy training systems at scale.

Keywords: Speech-Based Training, Empathy Training, Design Science Research, Self-Regulated Learning.

1 Introduction

Empathy represents the ability to understand another's perspective and to relate to the other in an emotionally appropriate way (Davis, 1983). Already Stephen Hawking has described empathy as the most important interpersonal skill (Hawking, 2002), since it is essential for social communication between individuals (Elliott *et al.*, 2011; de Vignemont and Singer, 2006). Thus, empathy functions as an essential part of communication competence by increasing one's understanding of and adjustment to others (Redmond, 1989). One way for humans to express their emotions in communications is through speech since the voice has been shown as an essential influence on the perception of empathy to others. Thus, emotions are communicated through both the substance and the manner in which something is delivered (Kraus, 2017). In addition to the growing importance of emotional abilities as an essential 21st century skill (OECD, 2018), a meta-analysis has shown that the empathy of college students has declined in recent decades. From 1979 to 2009, the empathy of college students decreased by more than thirty percent. The study also indicates that this decrease was even stronger between 2000-2009 (Konrath *et al.*, 2011). Hence, researchers and educators are calling for empathy skill training to take a more important role in today's education curricula (Gerdes *et al.*, 2011). Empathy is traditionally trained through experiential learning scenarios such as communication trainings or role-plays (Lam *et al.*, 2011).

In the past, these methods are mainly used to train medical personnel or students of social sciences (Foster *et al.*, 2016). Much experience-based training positively affects students' emotional skills, suggesting that corresponding empathy training is effective (Peterson and Limbu, 2009; Teding Van Berkhout, 2015). Nevertheless, educational institutions and universities face the problem of increasingly frequent large-scale lectures and distance learning scenarios, such as massive open online courses (Maruping and Matook, 2020; Seaman *et al.*, 2018), where only a limited amount of interaction is possible. These challenges, which are more likely to increase in the future, make it more difficult to provide students with personal and experiential training for emotional skills such as empathy. Therefore, the question arises, of how students can effectively train empathy in spoken communications under organizational limitations. Different studies already show that empathy can be effectively trained using technology (Lok and Foster, 2019; Santos *et al.*, 2018; Wambsganss *et al.*, 2022). Accordingly, using technology-enhanced applications seems to be an excellent way to individually train skills such as empathy in large-scale lectures and distance learning scenarios (Wambsganss, Weber, *et al.*, 2021). The recognition of empathic structures and phrases is a growing field. Corresponding approaches can be found for textual empathy detection (Buechel *et al.*, 2018; Wambsganss, Niklaus, *et al.*, 2021) as well as for the analysis of empathy voice features in speech-based data (Alam *et al.*, 2017; Xiao *et al.*, 2014). With the growing advances in voice analytics, one solution to help students improve their empathy skills in verbal communication seems to be a technology-mediated training system. Therefore, we seek to answer the following research question:

RQ: *How to design an empathy training system to help students develop their empathy skills in verbal communication?*

To answer our research question, we started a design science research (**DSR**) endeavor following Hevner *et al.* (2004). We seek to develop design principles (**DP**) to help practitioners and researchers develop appropriate training systems in a structured way. In the theoretical background, we follow the self-regulation model according to Winne and Hadwin (1998), which focuses on active training and considers our developed system as an external supporting factor for the learner. To the best of our knowledge, there is no other research that addressed the development of a training system to increase empathy in spoken communications through active training. We, therefore, derive five initial DPs, implement them in a prototype, and evaluate this prototype in a proof of concept evaluation.

2 Theoretical Background

2.1 Emotion and Empathy Detection in Voice Analytics

In voice analytics, empathy recognition in speech data is usually considered a subfield of emotion recognition. According to voice analysis research approaches, the human voice can be fully described as sound waves in four dimensions: time, amplitude, frequency, and spectrum (Hildebrand *et al.*, 2020; Jurafsky and Martin, 2014; Sueur, 2018). The analysis of these different characteristics can reveal information about the state of a speaker (Hildebrand *et al.*, 2020). Thus, the way we speak reveals something about the emotional state we are in (Jurafsky and Martin, 2014; Scherer, 2003) and influences how others perceive us. For instance, anger is associated with louder speech (Clark, 2005) and fear with greater pitch variability (Juslin and Laukka, 2003). Following this reasoning, empathy can also be perceived in the voice via vocal cues. The field of empathy analysis is still relatively small, but already some work mentions certain voice features concerning empathy. The usual linguistic features are voice pitch, speaking rate, and volume (Apple *et al.*, 1979; Weiste and Peräkylä, 2014; Xiao *et al.*, 2015).

2.2 Self-regulated learning

Self-regulated learning (SRL) theory provides a foundation for understanding the cognitive, motivational, and emotional aspects of learning (Panadero, 2017). Research interest in the theory has increased significantly in recent decades, and several models have been developed to conceptualize it

(Sitzmann and Ely, 2011). In this project, we follow the model of Winne and Hadwin (1998), which is used in research to implement computer-based learning environments. SRL is considered a constructive process in which learners set learning goals and then seek to monitor, regulate, and control their cognitive and metacognitive abilities (Panadero *et al.*, 2016). Current scientific work assumes that the effectiveness of SRL processes can also be transferred to learning in computer-based environments or with intelligent tutoring systems (Greene and Azevedo, 2007; Winne and Baker, 2013). We assume that an appropriate system to foster learners' empathy skills can stimulate and support users' self-regulatory processes, leading to increased self-efficacy, motivation, and engagement in learning (Zimmerman and Schunk, 2001). Therefore, various interventions provided by the system, such as support or feedback, should be able to help learners realize the full potential of SRL.

3 Methodological Approach

This research is guided by the DSR approach, according to (Gregor and Jones, 2007; Hevner *et al.*, 2004). Figure 1 illustrates the three cycles of the DSR, showing the steps we perform (greyed), their sequence, and the corresponding subsections that contain additional detailed information.

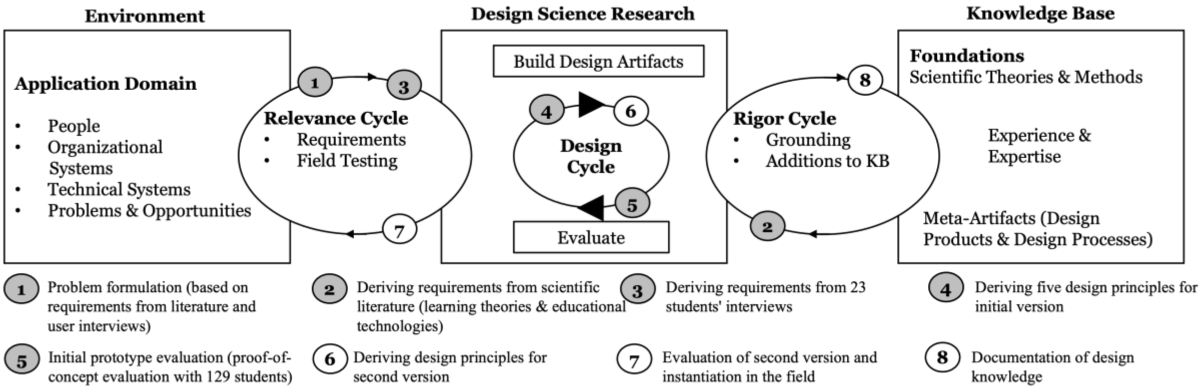

Figure 1. Three cycle design science process (Hevner et al., 2004; Gregor and Jones, 2007).

In this research in progress paper, we applied the first five steps of the DSR approach (Gregor and Hevner, 2013). The other three steps are to be applied in the further research project. Following the DSR approach, we have formulated the practical problem in the first step, which was comprehensively presented in the introduction. In the second step, we derived meta-requirements (**MR**) to be able to use the already existing knowledge base from the literature for the DPs. The literature was analyzed according to the systematic literature search procedure of Vom Brocke et al. (2015). In addition to the scientific knowledge base, we also consider the environment in which the system should be used. For this, we interviewed a total of 23 students (20 master's and three bachelor's students), following the structure of Gläser and Laudel (2009). To complement the MR from the scientific literature, user stories (**US**) and user requirements (**UR**) were derived based on the statements of the interviewed students. In the fourth step, five initial DPs were formed. Finally, in the fifth and last step, the initial prototype was evaluated to validate. An initial prototype was developed for this evaluation, and 129 students participated. Participants were recruited via the SurveyCycle platform, in lectures and student groups. The prototype was evaluated in a free simulation experiment (Gefen, 2000; Söllner *et al.*, 2016), in which participants solve predefined tasks but are free to experience the contents of the prototype (Söllner *et al.*, 2016). After the prototype test, the participants fill in a survey and give suggestions for improvement in free text.

4 Designing the Artefact

In the following, the derivation of the requirements and the initial DPs are presented (Table 1). For the formulation of the DPs, the structure of Gregor *et al.*, 2020 was used.

4.1 Deriving Meta-Requirements from Scientific Literature / User Interviews

The literature issues (**LI**) collected are composed of two scientific areas. First, literature was collected that deals with speech-based empathy training, secondly, we have collected theories from the field of education and technology-mediated learning. Davis' multidimensional approach **(LI1)** serves as the basis for the prototype empathy training system, thus the training content considers the emotional and cognitive empathy dimensions **(MR1)** (Davis, 1983; Wambsganss, Weber, *et al.*, 2021). In spoken communication, there are adaptation processes regarding the rhythm and the speed of speech **(LI2)** (Hatfield *et al.*, 2009). This emotional and vocal contagion, as well as the increased empathic accuracy through vocal communication (**LI3**; Kraus, 2017), lead to the focus on spoken communication and learners' voice features **(MR2)**. The aspects analyzed are based on the findings on vocal interpersonal communication **(LI4)** and therefore include the dimensions of voice pitch, speaking rate, and loudness **(MR3)** (Apple *et al.*, 1979; Weiste and Peräkylä, 2014; Xiao *et al.*, 2014). Regarding the learning process, it is shown that examples can reduce the cognitive load in learning (worked-example effect; **LI5**) (Sweller and Cooper, 2009). This reduction in the self-regulated learning process should be achieved as best as possible by providing learners with examples of empathic language-based communication **(MR4)**. Learning goals are important for constructive SRL processes. The positive effects of concrete goals on performance are also shown in goal-setting theory **(LI6)** (Locke and Latham, 2006), therefore training outcomes through goals and feedback on demonstrated learning performance are built into the system **(MR5)**.

Initial design principles (DPs)	**Requirements**		
For designers and researchers to design a speech-based empathy training system for students to increase their empathy skills based on vocal communication …	**MRs**	**LIs**	**URs**
DP1: …, they should provide input about the importance of empathic voice features in communication and scientific empathy definitions to facilitate a target-oriented learning process based on current scientific knowledge.	MR1,2	LI1	UR1
DP2: …, the system should include practical role-plays, which the user solves with a naturally spoken text to provide individualized feedback on the emphatic features in the users' voice.	MR1	LI2,3	UR2,3
DP3: …, they should employ individualized visual feedback on vocal cues and empathy dimensions (emotional and cognitive empathy) to enable practice-oriented training improvements regarding empathic communication.	MR3	LI4	UR4,5
DP4: …, they should include an auditive and verbal feedback channel limited to 30 to 60 seconds to provide examples of empathic communication.	MR4	LI5	UR2
DP5: …, they should integrate an overview of the training process to allow students to review and monitor their learning process by achieving learning goals.	MR5	LI6	UR6

Table 1. Presentation of the derived design principles according to Gregor et al., 2020.

To derive requirements from students as potential users of our speech-based empathy training system, we interviewed 23 students (20 master's and three bachelor's students), mainly from the fields of economics and social sciences. For these, we followed the instructions of Gläser and Laudel (2009). Of those interviewed, eight were male and 15 were female and the mean age was 25.09 years. A pre-structured questionnaire was used, which contained 25 questions and was divided into three parts. The following aspects were addressed: *the relevance of social and emotional skills in university education*, *considerations for the implementation of a technology-based training system*, and the *expectations/ requirements for a speech-based training system*. The interviews (mean: 42.82 minutes) were transcribed and coded. For this purpose, inductive coding (one round) was applied and the categories were derived accordingly from the transcribed material (Mayring, 2016). Based on these coded transcripts and the individual statements of the students, a total of 145 US were derived (Lucassen *et al.,* 2016). The most frequently mentioned and the most significant US were selected in the next step for formulating the UR. The derived URs confirm recent findings from the development of empathy training systems but also show new requirements to be explicitly considered for the development of speech-based training systems.

4.2 Initial Design Principles

The MR and UR identified in the preceding research formed the basis for the five initial DPs. The first refers to the theoretical framework in the empathy training system (**DP1**). This is implemented by providing users with scientific input on the empathic voice features in communication before the training task and the concluding feedback. This ensures that the user gets an overview of empathy and communication, as mentioned in the interviews conducted (**US1**) before it is applied in the training task (**UR1**). The training scenario is implemented through a practical role-play situation (**UR3**). We have chosen a training scenario that is easy to set up and relevant to possible real-life application situations for students. In the artifact, the user gives peer feedback with a spoken answer to a fellow student on a presented business pitch (**DP2**). This training task design should enable the provision of specific feedback on the empathic features in the users' voices (**UR4**). After the exercise has been completed, the user receives individualized visual feedback on the detected emotions in the answer (anger, sadness, happiness and fear), based on voice features (pitch rate, speaking rate and loudness) (Apple *et al.*, 1979; Weiste and Peräkylä, 2014; Xiao *et al.*, 2015) and on the empathy dimensions (Davis, 1983; Wambsganß *et al.*, 2021) (**DP3**). The dimensions of cognitive and emotional empathy arise from the empathy construct of Davis (1983). Cognitive empathy represents the ability to understand another person's perspective. While emotional empathy is described as the emotional response to another person's perspective. The visual design of the feedback is implemented in a dashboard, which presents the user an intuitive and individualized overview of the voice features and empathic dimensions (Wambsganss, Niklaus, *et al.*, 2021; Wambsganss, Weber, *et al.*, 2021). The feedback through which the user's learning progress is facilitated considers their demonstrated solution in the task and communication and empathy theory. The dashboard shows the expressions of different emotions detected in the response. In addition to the visual feedback, the system includes an auditive channel that provides verbal feedback (**DP4**). This verbal audio channel is intended to draw the user's attention to important elements and aspects of their spoken responses. Therefore, the user is shown in detail which responses can be considered empathetic in relation to the speech feature. It was repeatedly mentioned in the interviews that students would like to be able to track their progress through learning goals (**UR6**), which is consistent with the goal-setting theory (**MR5**). Medals in bronze, silver, and gold documented this progress in the system to motivate the users (**DP5**).

4.3 Initial Proof of Concept Evaluation

In the next step, the DPs were implemented in an initial prototype of a speech-based empathy training system. The goal of the evaluation was to validate or falsify our principles, which we derived from the scientific literature and conducted expert interviews with students. The results can help to adapt the DPs or consider new suggestions for possible DPs through the participants' input. For the evaluation, we follow the structure of Venable *et al.* (2016). It is conducted as an ex-ante evaluation with a prototype to decide how and in which form the development of comparable technology- and language-based empathy training systems can be continued (Stefanou, 2017; Venable *et al.*, 2016).

As described earlier, the task for participants was to provide peer feedback on a presented business pitch in a conversation with the system (Figure 3). To enable a first evaluation of our initial prototype the participant could choose between three pre-defined vocal answers (Appendix, Table 5), the user was supposed to first give *positive feedback* and *suggestions for improvements* afterward. After completing the task, the users received their feedback, in particular on voice features. For the voice analysis, the prerecorded voice messages were evaluated with the program “Praat” (Boersma and Van Heuven, 2007) with respect to pitch, speaking rate and volume. The prerecorded options differed in distinctive voice features and were classified as less or more empathic following previous scientific work (Figure 2; Appendix, Table 4). It must be noted that the emotional dimension of empathy is more affected by the evaluation of voice features than the cognitive one. The cognitive component is more concerned with understanding the opponent's perspective and less with the adequate response (as is the case with the emotional component). The more detailed visualization for the users was done in the emotional states of fear, anger, happiness and sadness. Accordingly, the users received visual and auditory feedback through a voice message to what extent their spoken reaction seemed appropriate to the situation. Participants received no predefined instruction on how to use the prototype; instead, they could navigate through the system intuitively. A timer was used to ensure that the prototype was tested for an appropriate amount of time before the subsequent survey was completed. In addition, responses that could not answer the integrated control question were weeded out. A total of 129 valid responses from students (63 bachelors, 66 masters) could be evaluated. 91 female and 36 male students participated; 2 identified as diverse. The mean age of all participants was 24.71 years.

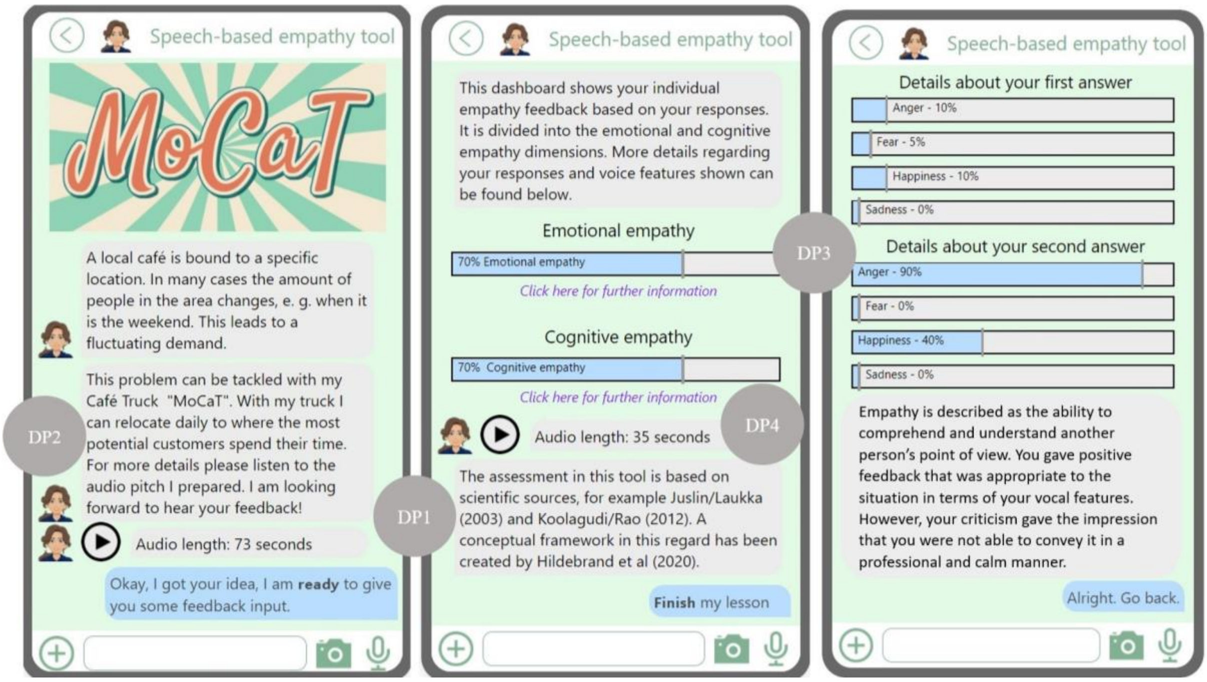

Figure 2. Initial prototype used for the evaluation.

In the post-test phase, the technology acceptance model of Agarwal and Karahanna, 2000; Venkatesh and Bala, 2008 was queried with the perceived ease of use (PEOU), perceived usefulness (PU) and intention to use (ITU). The items were adapted according to our training system. In addition, to be able to evaluate the system in terms of self-regulated learning, we queried the constructs "enjoyment" (Kim *et al.*, 2019; Lund, 2001; Shamekhi *et al.*, 2018), "cognitive load" (CL) (Krell, 2017; Lin *et al.*, 2020) and "intrinsic motivation" (Lin *et al.*, 2013, 2020). We used the constructs (5 points Likert scale) to identify if the DPs in the system can support self-regulatory processes, leading to higher self-efficacy and engagement in learning (Zimmerman and Schunk, 2001). Concluding, the DPs themselves were evaluated with the following questions or items: "*Receiving scientific input on the emotional and cognitive empathy dimensions is helpful for me*" (**DP1**), "*I think that speech-based practical scenarios would be useful for me to give better empathic feedback in the future*" (**DP2**), perceived feedback accuracy following Podsakoff and Farh (1989) (**DP3**) "*I find the possibility to listen to the feedback and therefore also to have an example of empathic communication useful in terms of giving empathic feedback*" (**DP4**) and "*The overview of the learning progress would motivate me to do more exercises in the tool*" (**DP5**).

5 Results

The evaluation of the formulated initial DPs showed promising results (Table 2). For this purpose, the mean values were calculated for a first tendency of the evaluation results and a Wilcoxon test was performed for the sample. The results show that the evaluated DPs scored significantly higher than the mean of 3 (based on the 5 points Likert scale). The scaled mean of 3 was chosen as a neutral comparison value since we are currently not aware of any other system with which we could have conducted a different control group evaluation. In particular, the inclusion of definitions and information about empathy and the voice features in communication (**DP1**: 3.8) and the subsequent practical task of giving peer feedback (**DP2**: 3.65) show that this set-up was positively perceived. The possibility of listening to the feedback auditorily in a voice message was also positively evaluated (**DP**4: 3.79). Overall, a significant difference from the mean was shown in four of the five initial DPs, which scored higher than the mean on average.

n=129	DP1	DP2	DP3	DP4	DP5
Mean	3.8	3.65	3.24	3.79	3.58
Median	4***	4***	3,33	4***	4***
z-value	7,525	5,965	2,639	6,912	5,678
p<=0.05*; p<=0.01**; p<=0.001***					

Table 2. Results of evaluation of DP 1 to 5 (scale: 5 points Likert).

Since the results regarding the further evaluated constructs did not show normally distributed data, a one-sample Wilcoxon signed-rank test was performed to make a first evaluation of the prototype. The results for perceived usefulness and perceived ease of use are also promising in the first analysis (Table 3). The fact that participants perceive the operation of the speech-based system as easy (mean: 4.14; median 4,25) and the training content as useful for giving empathic feedback (mean: 3.57; median 3,67) has a positive effect on the adaptation for later use in the sense of the TAM model. Another construct that positively influences the acceptance and possible adoption of new technology is enjoyment during use (Agarwal and Karahanna, 2000). Intrinsic motivation (mean: 3.52) was positively rated by the participants, while cognitive load (mean: 2.98; median 3,0) was approximately evaluated with the neutral reference value of 3.

n=129	Perceived usefulness	Perceived ease of use	Intention to use	Cognitive load	Enjoyment
Median	3,67***	4,25***	3,33	3,00	3,4***
z-value	6,528	9,384	1,672	0,576	4,175
p<=0.05*; p<=0.01**; p<=0.001***					

Table 3. Results of evaluation (scale: 5 points Likert).

The respondents also formulate qualitative feedback and make suggestions for improving the system. Participants emphasized in seven notes that they enjoyed using the prototype and that its usability was good: "*Easy handling, short and useful descriptions*." Additionally, students stated that there was a "*very close examination of my ans*wers." In essence, we see this as a positive perception of the feedback. In addition to the positive statements, we also want to explain the suggestions for improvement. Here, the focus is on the response options and the interaction design: "*I would have liked to give answers by myself, not just by clicking on the Options. But I like the idea and the structure of the learning tool*." The students noted that they would like to have more answer options to be able to give feedback: "*Not enough answer options to choose from, until it started only 3 answer options*". More crucially, students would like to voice responses themselves to receive individual feedback: "*There were only three options, and they were pre-voiced, so I feel as though my measure of empathy was inaccurate because, where I* *would've used similar voice inflections and intonation, my responses would've been different*." Both improvement or response possibilities and possibilities to give individual speech-based feedback were often mentioned. Nevertheless, it can be stated here that the goal is for users to record their voices and receive feedback in response. Some participants also stated that complexity is too high to concentrate on the content and voice cues. This should be mitigated if students can record their messages themselves, as they then no longer have to focus on predefined voice cues and content. Another idea of the participants was to enrich the system with videos to track gestures and facial expressions. The idea is interesting, but we assume this would go beyond the scope of the system. If the users had to concentrate on the content of their answers and on the language, facial expressions and gestures, the cognitive load would be too high, and the learning effect would be lost (Sweller *et al.*, 2011). All in all the first results of the evaluation of the initial prototype and the DPs are promising but still need to be consolidated. A further evaluation is planned with a comparable alternative system and a control group to support the preliminary results (Abbasi et al., 2010).

6 Conclusion and Outlook

This research in progress paper presents five initial DPs for developing a speech-based empathy training system and the first evaluation of an initial prototype. As a limitation of our work, we would like to state that the requirements for the system were derived from a specific view of the literature and with a specific user group. Therefore, other researchers might come to different results with a different literature strand and a different user group. In addition, it is important to mention that DP1, DP2 and DP5 also appear similarly in already established works for technology-based skill training (Rietsche *et al.*, 2018; Wambsganss *et al.*, 2020; Wambsganss, Weber, *et al.*, 2021). This is because users desire similar requirements for technology-based training systems and specific requirements are modus inverse. They are therefore relevant for speech-based and for textual training systems. However, we would like to clarify that all DPs have evolved from user interviews with students to developing a speech-based training system, thus the DPs are essential for developing appropriate systems. We want to continue our work and evaluate the initial DPs and the developed prototype further. To reach this, we would like to carry out the expanded steps 6 and 7 of the DSR according to Hevner (2007). Based on the first results of this study, a training system will be developed, which will be evaluated in large-scale lectures (Step 7; Figure 1). Thereby we hope to get further results about the effectiveness of a speech-based empathy training system.

Appendix

		Emotional state			
		Happiness	**Anger**	**Sadness**	**Fear**
Vocie features	**Pitch Rate**	Increasing ▲	Increasing ▲	Decreasing▼	Increasing ▲
	Speaking Rate	Faster ▲	Faster▲	Slower ▼	Faster ▲
	Loudness	Increasing▲	Increasing ▲	Decreasing ▼	Decreasing▼

Table 4. Voice features and emotional states (Hildebrand *et al.*, 2020; Juslin and Laukka, 2003).

Answer Options		**Pitch Rate**	**Loudness**	**Speaking Rate**
Neutral Reference		137,6 Hz	65,9 dB	4,4 syllables/Sec
Question 1	Option 1.1	139,4 Hz	67,6 dB	4,4 syllables/Sec
	Option 1.2	178,3 Hz	76,2 dB	5,1 syllables/Sec
	Option 1.3	132,8 Hz	59,1 dB	3,4 syllables/Sec
Question 2	Option 2.1	178,3 Hz	76,2 dB	5,3 syllables/Sec
	Option 2.2	137,2 Hz	65,4 dB	4,5 syllables/Sec
	Option 2.3	133,1 Hz	56,8 dB	3,5 syllables/Sec

Table 5. Presentation of the voice analysis for the response options.

References

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H. and Nunamaker, J.F. (2010), “Detecting fake websites: The contribution of statistical learning theory”, *MIS Quarterly: Management Information Systems*, University of Minnesota, Vol. 34 No. SPEC. ISSUE 3, pp. 435–461.

Agarwal, R. and Karahanna, E. (2000), “Time flies when you’re having fun: Cognitive absorption and beliefs about information technology usage”, *MIS Quarterly: Management Information Systems*, Management Information Systems Research Center, Vol. 24 No. 4, pp. 665–694.

Alam, F., Danieli, M. and Riccardi, G. (2017), “Annotating and modeling empathy in spoken conversations”, *Computer Speech & Language*, Vol. 50, pp. 40–61.

Apple, W., Streeter, L.A. and Krauss, R.M. (1979), “Effects of pitch and speech rate on personal attributions”, *Journal of Personality and Social Psychology*, American Psychological Association Inc., Vol. 37 No. 5, pp. 715–727.

Boersma, P. and Van Heuven, V. (2007), “Speak and unSpeak with PRAAT”, *Glot International*, Vol. 5 No. 9/10, pp. 341–347.

Vom Brocke, J., Riemer, K. and Niehaves, B. (2015), “Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research”, *Communications of the Association for Information Systems*, Vol. 37 No. 1, pp. 205–224.

Buechel, S., Buffone, A., Slaff, B., Ungar, L. and Sedoc, J. (2018), “Modeling Empathy and Distress in Reaction to News Stories”, *Conference on Empirical Methods in Natural Language Processing*, pp. 4758–4765.

Clark, A. (2005), *Psychology of Moods*, Nova Publishers, New York.

Davis, M.H. (1983), “Measuring individual differences in empathy: Evidence for a multidimensional approach”, *Journal of Personality and Social Psychology*, American Psychological Association Inc., Vol. 44 No. 1, pp. 113–126.

Elliott, R., Bohart, A.C., Watson, J.C. and Greenberg, L.S. (2011), “Empathy”, *Psychotherapy*, Vol. 48 No. 1, pp. 43–49.

Foster, A., Chaudhary, N., Kim, T., Waller, J.L., Wong, J., Borish, M., Cordar, A., *et al.* (2016), “Using Virtual Patients to Teach Empathy”, *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, Lippincott Williams and Wilkins, Vol. 11 No. 3, pp. 181–189.

Gefen, D. (2000), “E-commerce: the role of familiarity and trust”, *Omega*, Pergamon, Vol. 28 No. 6, pp. 725–737.

Gerdes, K.E., Segal, E.A., Jackson, K.F. and Mullins, J.L. (2011), “Teaching empathy: A framework rooted in social cognitive neuroscience and social justice”, *Journal of Social Work Education*, Routledge , Vol. 47 No. 1, pp. 109–131.

Gläser, J. and Laudel, G. (2009), *Experteninterviews Und Qualitative Inhaltsanalyse. 3., Überarbeitete Auflage*, 3rd ed.

Greene, J. and Azevedo, R. (2007), “A theoretical review of Winne and Hadwin’s model of self-regulated learning: New perspectives and directions”, *Review of Educational Research*, Vol. 77 No. 3, pp. 334–372.

Gregor, S., Chandra Kruse, L., Seidel, S. and Kruse, C. (2020), “The Anatomy of a Design Principle Design Knowledge Formulation and Reuse View project Capital, Systems, and Objects View project”, *Article in Journal of the Association for Information Systems*, pp. 1–49.

Gregor, S. and Hevner, A.R. (2013), “Positioning and presenting design science research for maximum impact”, *MIS Quarterly*, Vol. 37 No. 2, pp. 337–355.

Gregor, S. and Jones, D. (2007), “The anatomy of a design theory”, *Journal of the Association for Information Systems*, Association for Information Systems, Vol. 8 No. 5, pp. 312–335.

Hatfield, E., Rapson, R. and Le, Y. (2009), “Emotional contagion and empathy”, in Decety, J. and Ickes, W. (Eds.), *Books.Google.Com*, MIT Press, Cambridge, pp. 19–30.

Hawking, S. (2002), *The Theory of Everything: The Origin and Fate of the Universe*, New Millennium Press., Beverly Hills.

Hevner, A.R. (2007), “A Three Cycle View of Design Science Research”, *Scandinavian Journal of Information Systems*, Vol. 19 No. 2.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004), "Design science in information systems research", *MIS Quarterly: Management Information Systems*, Management Information Systems Research Center, Vol. 28 No. 1, pp. 75–105.

Hildebrand, C., Efthymiou, F., Busquet, F., Hampton, W.H., Hoffman, D.L. and Novak, T.P. (2020), "Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications", *Journal of Business Research*, Elsevier, Vol. 121, pp. 364–374.

Jurafsky, D. and Martin, J. (2014), *Speech and Language Processing*, *Speech and Language Processing*, Vol. 3, Pearson, London.

Juslin, P. and Laukka, P. (2003), "Communication of emotions in vocal expression and music performance: Different channels, same code?", *Psychological Bulletin*, Vol. 5 No. 129, pp. 770–814.

Kim, S., Lee, J. and Gweon, G. (2019), "Comparing Data from Chatbot and Web Surveys: Effects of Platform and Con-versational Style on Survey Response Quality", *ACM Reference Format: Soomin Kim*, ACM, pp. 1–12.

Konrath, S.H., O'Brien, E.H. and Hsing, C. (2011), "Changes in dispositional empathy in American college students over time: A meta-analysis", *Personality and Social Psychology Review*, SAGE PublicationsSage CA: Los Angeles, CA, Vol. 15 No. 2, pp. 180–198.

Kraus, M.W. (2017), "Voice-only communication enhances empathic accuracy.", *American Psychologist*, Vol. 72 No. 7, pp. 644–654.

Krell, M. (2017), "Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence", edited by Hui, S.K.F.*Cogent Education*, Taylor and Francis Ltd., Vol. 4 No. 1, pp. 1–10.

Lam, T., Kolomitro, K. and Alamparambil, F.C. (2011), "Empathy training: Methods, evaluation practices, and validity", *Journal of Multidisciplinary Evaluation*, Vol. 7 No. 16, pp. 162–200.

Lin, L., Atkinson, R.K., Christopherson, R.M., Joseph, S.S. and Harrison, C.J. (2013), "Animated agents and learning: Does the type of verbal feedback they provide matter?", *Computers and Education*, Pergamon, Vol. 67, pp. 239–249.

Lin, L., Ginns, P., Wang, T. and Zhang, P. (2020), "Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain?", *Computers and Education*, Elsevier Ltd, Vol. 143, pp. 1–11.

Locke, E.A. and Latham, G.P. (2006), "New directions in goal-setting theory", *Current Directions in Psychological Science*, SAGE Publications Inc., Vol. 15 No. 5, pp. 265–268.

Lok, B. and Foster, A.E. (2019), "Can Virtual Humans Teach Empathy?", *Teaching Empathy in Healthcare*, Springer International Publishing, Cham, pp. 143–163.

Lucassen, G., Dalpiaz, F., van der Werf, J.M.E.M. and Brinkkemper, S. (2016), "Improving agile requirements: the Quality User Story framework and tool", *Requirements Engineering*, Springer London, Vol. 21 No. 3, pp. 383–403.

Lund, A.M. (2001), "Measuring usability with the use questionnaire12", *Usability Interface*, Vol. 8 No. 2, pp. 3–6.

Maruping, L.M. and Matook, S. (2020), "The evolution of software development orchestration: current state and an agenda for future research", *Https://Doi.Org/10.1080/0960085X.2020.1831834*, Taylor & Francis, Vol. 29 No. 5, pp. 443–457.

Mayring, P. (2016), *Qualitative Inhaltsanalyse: Grundlagen Und Techniken*, Vol. 12, Beltz Verlagsgruppe, Weinheim Basel.

OECD. (2018), "The future of education and skills: Education 2030", available at: https://www.voced.edu.au/content/ngv:79286 (accessed 17 November 2021).

Panadero, E. (2017), "A review of self-regulated learning: Six models and four directions for research", *Frontiers in Psychology*, Frontiers Media S.A., Vol. 8 No. APR, pp. 1–28.

Panadero, E., Klug, J. and Järvelä, S. (2016), "Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand", *Scandinavian Journal of Educational Research*, Routledge, Vol. 60 No. 6, pp. 723–735.

Peterson, R.T. and Limbu, Y. (2009), "The convergence of mirroring and empathy: Communications training in business-to-business personal selling persuasion efforts", *Journal of Business-to-Business Marketing*, Vol. 16 No. 3, pp. 193–219.

Podsakoff, P.M. and Farh, J.L. (1989), "Effects of feedback sign and credibility on goal setting and task performance", *Organizational Behavior and Human Decision Processes*, Academic Press, Vol. 44 No. 1, pp. 45–67.

Redmond, M. V. (1989), "The Functions of Empathy (Decentering) in Human Relations", *Human Relations*, Vol. 42 No. 7, pp. 593–605.

Rietsche, R., Duss, K., Persch, J.M. and Söllner, M. (2018), "Design and Evaluation of an IT-based Formative Feedback Tool to Foster Student Performance", *N International Conference on Information Systems (ICIS)*, pp. 1–17.

Santos, B.S., Junior, M.C. and De Souza, J.G. (2018), "An Experimental Evaluation of the NeuroMessenger: A Collaborative Tool to Improve the Empathy of Text Interactions", *Proceedings - IEEE Symposium on Computers and Communications*, Vol. 2018-June, Institute of Electrical and Electronics Engineers Inc., pp. 573–579.

Scherer, K.R. (2003), "Vocal communication of emotion: A review of research paradigms", *Speech Communication*, North-Holland, Vol. 40 No. 1–2, pp. 227–256.

Seaman, J., Allen, I. and Seaman, J. (2018), *Grade Increase: Tracking Distance Education in the United States.*, *ERIC*, available at: https://eric.ed.gov/?id=ED580852 (accessed 17 November 2021).

Shamekhi, A., Liao, Q.V., Wang, D., Bellamy, R.K.E. and Erickson, T. (2018), "Face value? Exploring the effects of embodiment for a group facilitation agent", *CHI Conference on Human Factors in Computing Systems* , Association for Computing Machinery, pp. 1–13.

Sitzmann, T. and Ely, K. (2011), "A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go.", *Psychological Bulletin*, Vol. 137 No. 3, pp. 421–442.

Söllner, M., Hoffmann, A. and Leimeister, J.M. (2016), "Why different trust relationships matter for information systems users", *European Journal of Information Systems*, Taylor & Francis, Vol. 25 No. 3, pp. 274–287.

Stefanou, C.J. (2017), "A framework for the ex-ante evaluation of ERP software", *European Journal of Information Systems* , Taylor & Francis, Vol. 10 No. 4, pp. 204–215.

Sueur, J. (2018), *Sound Analysis and Synthesis with R*, Springer International Publishing, Cham, available at:https://doi.org/10.1007/978-3-319-77647-7.

Sweller, J., Ayres, P., Kalyuga, S., Sweller, J., Ayres, P. and Kalyuga, S. (2011), "Measuring Cognitive Load", *Cognitive Load Theory*, Springer New York, pp. 71–85.

Sweller, J. and Cooper, G.A. (2009), "The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra", *Http://Dx.Doi.Org/10.1207/S1532690xci0201_3*, Lawrence Erlbaum Associates, Inc. , Vol. 2 No. 1, pp. 59–89.

Teding Van Berkhout, E. (2015), "The Efficacy of Empathy Training: A Meta-Analysis of Randomized Controlled Trials", *Journal of Counseling Psychology*, Vol. 63 No. 1, pp. 32–41.

Venable, J., Pries-Heje, J. and Baskerville, R. (2016), "FEDS: a Framework for Evaluation in Design Science Research", *Orsociety.Tandfonline.Com*, Palgrave Macmillan Ltd., Vol. 25 No. 1, pp. 77–89.

Venkatesh, V. and Bala, H. (2008), "Technology acceptance model 3 and a research agenda on interventions", *Decision Sciences*, Vol. 39 No. 2, pp. 273–315.

de Vignemont, F. and Singer, T. (2006), "The empathic brain: how, when and why?", *Trends in Cognitive Sciences*, Vol. 10 No. 10, pp. 435–441.

Wambsganß, T., Guggisberg, S. and Söllner, M. (2021), "ArgueBot: A Conversational Agent for Adaptive Argumentation Feedback Understanding and Designing Trust in Information Systems View project Trustworthy Conversational AI (TCAI) View project", *Wirtschaftsinformatik* , pp. 1–18.

Wambsganss, T., Niklaus, C., Söllner, M., Handschuh, S. and Leimeister, J.M. (2021), "Supporting Cognitive and Emotional Empathic Writing of Students", *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 4063–4077.

Wambsganss, T., Söllner, M., Koedinger, K. and Leimeister, J.M. (2022), "Adaptive Empathy Learning Support in Peer Review Scenarios", *ACM CHI Conference on Human Factors in Computing System (CHI). - New Orleans, Louisiana, USA.* , available at: https://www.alexandria.unisg.ch/265977 (accessed 23 March 2022).

Wambsganss, T., Söllner, M. and Leimeister, J. (2020), "Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills", *International Conference on Information Systems (ICIS)*, pp. 1–17.

Wambsganss, T., Weber, F. and Söllner, M. (2021), "Design and Evaluation of an Adaptive Empathy Learning Tool", *Proceedings of the 54th Hawaii International Conference on System Sciences*, Hawaii International Conference on System Sciences, pp. 54–63.

Weiste, E. and Peräkylä, A. (2014), "Prosody and empathic communication in psychotherapy interaction", *Psychotherapy Research*, Routledge, Vol. 24 No. 6, pp. 687–701.

Winne, P. and Baker, R. (2013), "The potentials of educational data mining for researching metacognition, motivation and self-regulated learning", *Journal of Educational Data Mining*, Vol. 5 No. 1, pp. 1–8.

Winne, P.H. and Hadwin, A.F. (1998), "Studying as self-regulated engagement in learning", in Schunk, D.H. and Zimmermann, B.J. (Eds.), *Metacognition in Educational Theory and Practice*, Lawrence Erlbaum Associates, New York, pp. 277–304.

Xiao, B., Bone, D., Segbroeck, M., Imel, Z., Atkins, D., Georgius, P. and Narayanan, S. (2014), "Modeling therapist empathy through prosody in drug addiction counseling", *Fifteenth Annual Conference of the International Speech Communication Association*, pp. 213–2017.

Xiao, B., Imel, Z.E., Atkins, D.C., Georgiou, P.G. and Narayanan, S.S. (2015), "Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling", *Sixteenth Annual Conference of the International Speech Communication Association*, pp. 2489–2493.

Zimmerman, B. and Schunk, D. (2001), *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd ed., Erlbaum, Mawah New Jersey.