

An Exploratory Survey Questionnaire to Understand What Emotions Are Important and Difficult to Communicate for People with Dysarthria and Their Methodology of Communicating

Lubna Alhinti, Heidi Christensen, Stuart Cunningham

Abstract—People with speech disorders may rely on augmentative and alternative communication (AAC) technologies to help them communicate. However, the limitations of the current AAC technologies act as barriers to the optimal use of these technologies in daily communication settings. The ability to communicate effectively relies on a number of factors that are not limited to the intelligibility of the spoken words. In fact, non-verbal cues play a critical role in the correct comprehension of messages and having to rely on verbal communication only, as is the case with current AAC technology, may contribute to problems in communication. This is especially true for people's ability to express their feelings and emotions, which are communicated to a large part through non-verbal cues. This paper focuses on understanding more about the non-verbal communication ability of people with dysarthria, with the overarching aim of this research being to improve AAC technology by allowing people with dysarthria to better communicate emotions. Preliminary survey results are presented that give an understanding of how people with dysarthria convey emotions, what emotions that are important for them to get across, what emotions that are difficult for them to convey, and whether there is a difference in communicating emotions when speaking to familiar versus unfamiliar people.

Keywords—Alternative and augmentative communication technology, dysarthria, speech emotion recognition, VIVOCA.

I. INTRODUCTION

THE quality of a person's life is determined by multiple factors, and one of the fundamental ones is the quality of communication. People need to communicate to express their feelings and needs, share their thoughts, ask questions, socialize, etc. Spoken language is one of the most common ways to communicate and it plays a critical role in defining who we are. However, some people may lose or be born without an ability to communicate effectively. For example, many neuro-motor conditions, like cerebral palsy, can affect a person's articulatory system to such a degree that people not familiar with their speech pattern will struggle to understand them.

According to [1], the most common acquired speech disorder is dysarthria, which can be defined as a neurological disorder that affects different aspects of speech production

Lubna Alhinti is with the Department of Computer Science, University of Sheffield, United Kingdom (corresponding author, e-mail: laalhinti1@sheffield.ac.uk).

Heidi Christensen is with the Department of Computer Science, University of Sheffield, United Kingdom (e-mail: heidi.christensen@sheffield.ac.uk).

Stuart Cunningham is with the Health Sciences School, University of Sheffield, United Kingdom (e-mail: s.cunningham@sheffield.ac.uk).

caused by weakness in the muscles responsible for speaking, miscoordination or inaccuracy of articulatory movements, or irregularity in the tone, steadiness, or speed [2]. People with dysarthria face many barriers to communicate effectively and it is widely acknowledged that their lives can be impacted negatively. The impact heavily depends on the individual and on the severity of the dysarthria. In an exploratory study conducted by [1] on the speakers' experience living with acquired chronic dysarthria, six dimensions were illustrated where dysarthria has a negative influence on their lives. One of these dimensions is that having dysarthria changes their way of communicating. These changes were reflected in, but not limited to, the speaker's style and communication behavior, and the capacity of putting feelings into their voices. The speech of people with dysarthria can often be monotone due to their reduced ability to control the vibration of their vocal folds. Revealing emotions such as gratitude, enthusiasm, and anger through their voices in a way that could be understood by others can be difficult.

Human-to-human communication can be simply viewed as the process of producing and receiving messages. These messages are formulated using different signs and codes that are interpreted by the receiver [3]. With the emergence of Augmentative and Alternative Communication (AAC), people with speech disabilities have been given a way to support their communication. AAC technology users have new social roles and therefore have new desires, demands and expectations in their social participation that they would like to fulfill [4]. However, current AAC technologies do have its challenges in addressing all the needed communication interactions. One of the most reported challenges is their slow rate of communication [5]. Other challenges include:

- The lack of context-related vocabulary that supports ongoing communication activity.
- The inability to support different circumstances and environmental challenges due to the limited number of available input and output channels.
- The inability to construct pragmatically appropriate utterances for speech output as a way to increase the perceptual access of the technology. For example the single phrase "oh.", "oh?", and "oh!" [5], [6].

The limitations of the current AAC technology also negatively affect social interactions and relationships as its

design and function does not take into account how human factors affect human-to-human communication in different conditions and scenarios [6]. The current state-of-the-art of the AAC technology is still far from being ideal and its limitations stand as barriers to the optimal use of these technologies in daily communication routines.

The Voice Input Voice Output Communication Aid (VIVOCA) is a new AAC technology that is operated by voice. A VIVOCA can recognize and interpret a disordered speech utterance and produce the required message using synthetic speech. This technology has been evaluated on highly disordered speech and achieved high recognition results [7]. The authors in [5] and [6] have listed a set of recommendations for the development of an AAC technology that would allow fuller participation in society. In addition to their recommendations, we believe that an AAC device should be able to reflect the emotional state of its user as emotions play a critical role in establishing and maintaining relationships.

Dysarthric speech has been characterized prosodically by monopitch, monoloudness, abnormal speaking rate, F0 range and flexibility impairments, and/or vocal intensity deviation [8]. This can result in unintelligible speech. As a result, people often presume that people with dysarthria have cognitive problems, which is incorrect most of the times. People with dysarthria do, in general, not have any problems understanding other people's speech and emotions. Their communication problem is mainly about producing intelligible speech [9]. In addition, their way of conveying emotions may also be different from that of typical speakers. Having an unintelligible and monotonous voice can increase the potential of being socially withdrawn. Nevertheless, people with dysarthria show strong preference of using their residual voices when they communicate as it is the natural mean of communication [10]. The survey reported in this paper tries to answer some of the initial questions around how people with dysarthria manage to convey emotions and their preferences as to what emotions AAC should be able to convey.

People have been looking into different ways of applying expressiveness to synthetic speech. The study of the 17 ways to say "yes" revealed four perspectives of the voice tone: emotional state, conversational intent, social context, and vocal qualities [11]. Communicating emotions is part of the expressiveness that can be added to the AAC devices. There are several possible input channels that can be used to communicate emotions using AAC devices. For example, the use of emotion words from the AAC vocabulary list such as happy, sad, etc., or the use of visual emotional symbols. Out of the many possible ways, it would be interesting to be able to communicate emotions directly using VIVOCA especially as many AAC users prefer to use their residual voices. Fig. 1 presents a high level description of our hypothesized dysarthric speech driven AAC device in which the AAC device would be able to recognize and interpret its user's disordered speech along with their emotional state and then deliver the message with the effect of the detected emotion in a clear synthesized voice.

In this research, we aim to improve AAC technology, that is

used by people with dysarthria, to enable them to communicate non-verbal information such as emotion. This paper presents the results of an initial study towards this goal where we investigate how people with dysarthria convey emotions in their speech. We plan to use this information to help people with dysarthria to convey emotions through VIVOCA devices. We need to first establish an understanding of how, and to what extent, people with dysarthria are able to convey emotions. What are the channels they tend to use? What are the important emotions for them to get across in daily communication? What are the emotions that are difficult for them to communicate, and is there a difference in how emotions are communicated to familiar versus unfamiliar people? This paper presents preliminary results from a survey that covers the above questions. Establishing a good understanding of these findings is very important before we start the process of automating dysarthric speech emotion recognition. The rest of the paper is structured as follows. Section II describes paralinguistic information in dysarthric speech. Section III discusses the different models of emotions in speech emotion recognition and speech synthesis. Section IV describes the methodology followed in conducting this research. Section V presents the survey's preliminary results and findings. Finally, Section VI combines our conclusion and outline of future work.

II. PARALINGUISTIC INFORMATION IN DYSARTHIC SPEECH

The human voice is a powerful and complex mean of communication. Its power comes from the fact that it transmits many signals and information beyond spoken words. Biological, psychological, social and emotional status as well as age, sex and weight are examples of the information that could be transmitted through the voice [12]. Pitch, sound pressure, timber, and tone are the main characteristics of the human voice [13]. Emotions affect the way we speak by changing the speech pattern and tonal quality. For example, the use of shrill or high pitched voices may relate to a scared or panicked emotional state; the use of long pauses and slow rate of speaking may indicate a pensive emotional state and the use of lower intensity may indicate a sad or ashamed emotional state [13], [14].

Dysarthria interferes with articulation, respiration, phonation, and resonance. Therefore, dysarthria changes the way people communicate. Having limited phonological and prosody dimensions, may not only result in producing less intelligible speech, but it may make it harder to convey emotions in a way that can be understood clearly and easily by recipients. There has been a number of studies that investigated the paralinguistic, precisely acoustic, prosodic and phonatory features of dysarthric vocalization. The author in [15] conducted an experiment with eight speakers, with severe spastic dysarthria caused by cerebral palsy, to investigate the use of pitch contour and syllable duration for phrase-level productions and whether there exists a vocal control to signal a linguistic contrast. The speakers were asked to produce ten unique phrases. Each phrase was produced five times as a statement and five times as a question.

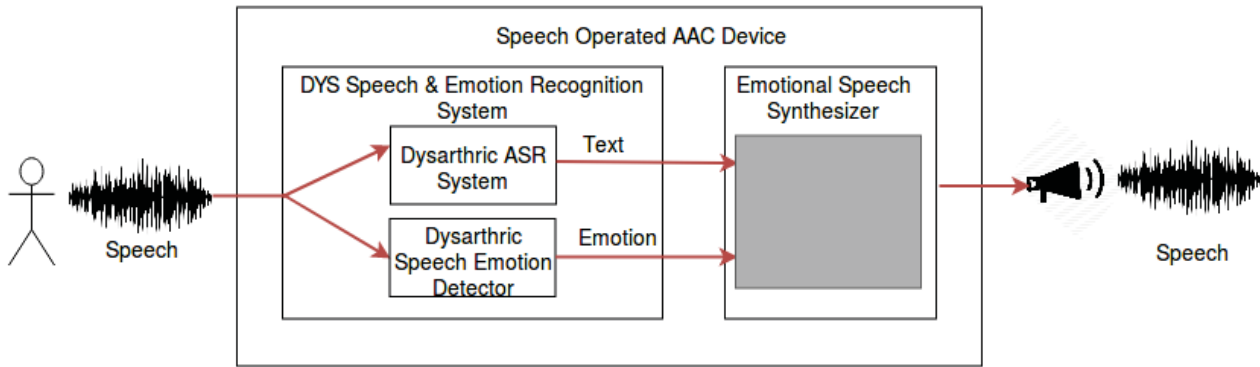


Fig. 1 Speech operated AAC device for people with dysarthria

For the evaluation, forty-eight normal-hearing English monolingual speakers who were unfamiliar with dysarthric speech and did not know the stimulus materials classified the production of statements and questions. Using prosodic cues only, the listeners were able to achieve approximately 87% classification accuracy. In the same study, F0 and duration cues were systematically removed to determine their importance in classifying dysarthric utterances. The effect of removing the pitch contour reduced listeners' accuracy scores by 32% while removing the durational cues reduced the accuracy by 8%. This implies that syllable duration is less informative than F0 contour to the listeners. In a follow-up study, the author in [8] replicated the above experiment with eight healthy controls to investigate the strategies used by the speakers with dysarthria due to cerebral palsy to signal the question-statement contrast, precisely, to find out what are the acoustic cues they use to do so and whether they are using different strategies to signal the contrast compared to the healthy controls speakers. The results indicate that F0, duration, and intensity were all used by speakers with dysarthria to signal the question-statement contrast while F0 and duration cues were primarily used by healthy control speakers. The author in [16] also examined the ability of people with severe dysarthria caused by cerebral palsy to convey information using pitch and duration cues. The study was conducted on a group of eight speakers with severe dysarthria caused by cerebral palsy. The results indicated that during sustained vowel production, all the eight speakers had consistent control over duration while producing the vowel /a/. They were all able to produce at least three different levels of vowel duration. Their ability to control the pitch, however, vary. All speakers were able to produce at least two different levels of F0.

Although the above studies were carried out with only eight speakers with severe dysarthria caused by cerebral palsy, and it is unclear how the results may generalize, it does however highlight several important points. It shows that these people may have enough control to communicate intentions, gain attention, and convey emotions [8], [15], [16]. This potential control of phonatory and prosodic features may open new doors to improve communication aids in a way that makes

it more sensitive to specific cues in the vocalization signal produced by the speaker with dysarthria and act according to the speaker's intention.

With the insight of the above studies and results, we believe that the dysarthric speech automatic emotion classification is a promising area and with more research, people with dysarthria may start communicating emotions more efficiently through technology. Since dysarthria affects the voice and limits the phonological and prosodic dimensions, we need to identify how people with dysarthria communicate emotions, whether people with dysarthria share the same acoustic features while conveying a certain state of emotion, and study the inter- and intra- speaker variations. The survey discussed in Sections IV and V is the first step in this research program with the eventual aim of automating the emotion recognition process.

III. MODELING EMOTIONS IN SYNTHETIC SPEECH

In recent years, research on Automatic Emotion Recognition (AER) and emotional speech synthesis have gained more attention due to the prospect of the substantial applications. There are two ways to conceptualize emotions. It is either done by using a discrete (categorical) approach, or a dimensional approach. In the discrete approach emotions are identified using a small number of basic and primary emotions. Usually six basic emotions are used: happiness, sadness, fear, surprise, anger, and disgust [17]. In contrast, in the dimensional approach, emotions are identified in a two dimensional model: valence and arousal [18]. The discrete emotion approach has been popularized by scholars in this field and many emotional speech databases have been developed following this approach. The Geneva Airport Lost Luggage Study [19], the Danish emotional speech database [20], the Berlin database of emotional speech [21], the eNTERFACE'05 audio-visual emotion database [22], and the Surrey Audio-Visual Expressed Emotion (SAVEE) database [23], to name a few examples of databases that adopted this approach in the development of their databases of *typical* speech.

The approach is well developed and progress with typical speech has been encouraging, however, we cannot follow the same approach directly on disordered speech. We first need to understand how people with disordered speech communicate

emotions: What kind of emotions are they capable of communicating? How well are these emotions recognized? What model is it best to followed when applying automatic speech emotion recognition: discrete or dimensional? The main aim of the survey discussed in Section IV and V is to establish the first step towards such an understanding and to quantify what are the more needed emotions to focus on initially.

IV. METHODOLOGY

This research received ethical approval from the ethical review panel of the Department of Computer Science at the University of Sheffield. It is composed of a number of stages. The aim of the first stage of this research is to achieve a better understanding of how people with dysarthria communicate emotions. Therefore, a survey was designed to address the following questions:

- How difficult it is for people with dysarthria to get their emotions across?
- What are the emotions that are important to them to get across?
- What are the ways that they tend to use to get their emotions across?
- Is there a difference in the way emotions are communicated to familiar and unfamiliar people?

Knowing the answers to the above questions will help in defining the scope of the research. It will also help in identifying the generalizability of this research area among people with dysarthria. The survey was distributed using special email lists that targeted participants who have dysarthria within the United Kingdom. The following section discusses and analyzes the main findings of the survey. The second stage of the research will deal with the technical part where we will investigate how different emotions can be automatically recognized from a dysarthric speech, how well these emotions can be recognized, and the effect of the intra- and inter- speaker variations. To facilitate that stage, we plan to record a database with parallel recording of emotional dysarthric and typical speech.

V. SURVEY RESULTS

The survey contains a total of 27 questions. Closed questions, open-ended questions and rank order questions were included to get the maximum information out of this survey. To follow a logical flow of the questions, the survey is arranged into three sections. The first section is related to the use of a communication aid. The second section, which is the main section, is related to questions about emotions. The third and final section is related to demographic information. Below, we will discuss preliminary findings of the main questions in the survey from eight respondents – six male, one female, and one participant who preferred not to say. Five of the respondents have severe dysarthria and three of them have moderate dysarthria.

All but one of the respondents are users of one form or another of a voice output communication aid. The non-communication aid user respondent indicated their

preference of using their residual speech over communication aids as it is a faster means of communication. The survey shows that people with dysarthria can face difficulties when communicating with familiar people, if they are not using their communication aid, but that this problem is exacerbated when they are communicating with unfamiliar people.

Given a list of seven different emotions (Happiness, Sadness, Anger, Surprise, Boredom, Disgust, and Fear), respondents were asked about what emotion do they feel is the most useful to try to communicate in their social life. All but one of the respondents chose Happiness, with the remaining respondent choosing Fear. *"People need to know that I am happy with them so they want to come back and be with me"*, a justification of the importance of communicating happiness in social life settings from one of the respondents. Happiness was also chosen by the majority of the respondents as the most important emotion they feel they want to communicate in everyday life.

From the set of questions that focus on addressing the difficulty of communicating emotions, the following question was asked: *"What emotion do you feel is the most difficult for you to communicate to familiar people?"*. Anger was chosen by almost half of the respondents. Surprise, Boredom, and Disgust were chosen by the other respondents. When communicating with unfamiliar people, respondents' answers vary. However, Anger and Boredom were the most chosen emotions among the others. These emotions, in typical speech, are perhaps characterized by being more subtle (Boredom) or easily confusable (Anger/Surprise) compared to e.g., Happiness [24], [25].

Looking into the channels that people with dysarthria tend to use when communicating emotions to familiar people, the following question was asked: *"How do you communicate your emotions to familiar people?"*. The majority of the respondents indicated their use of facial expressions and/or speech. The use of gestures, and eye gaze were also indicated by some respondents. There was little difference in respondent's answers to this question regardless of whether they are communicating with familiar or unfamiliar people.

In a ranking order question, respondents were asked to rank a set of emotions according to their importance to them in terms of being able to communicate them successfully *"For 1 being the most important and 7 being the least important, please number the following emotions according to their importance to you in terms of being able to communicate them successfully"*. Fig. 2 presents the results obtained from this question where the average ranking of each answer choice is displayed. The results show an indication of the importance of successfully communicating Happiness, Anger, and Sadness.

The survey is currently live and collecting more respondents which will enable us to generalize beyond the relatively small number of responses we currently have. The aim of this paper is to provide an insight and understanding of some related aspects to communicating emotions for people with dysarthria rather than give final conclusions at this stage. It can be inferred, however, how complex the problem is as many factors have already been indicated to playing a critical role in the way emotions are communicated. The nature of the person and the

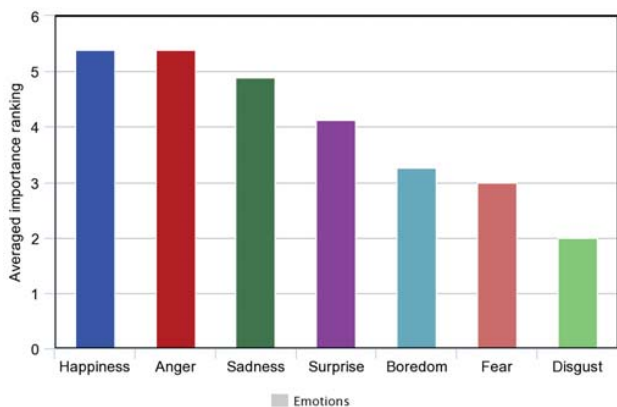


Fig. 2 Survey result of the most important emotions to communicate for people with dysarthria.

severity of their speech disorder are some of the main factors that may influence the way emotions are communicated.

VI. CONCLUSION

The survey shows that people with dysarthria find difficulties when communicating emotions with familiar people; however, the problem is exacerbated when they are communicating with unfamiliar people. Therefore perhaps a VIVOCA that could assist with this could be beneficial. This new field of research will need a lot of understanding of how people with dysarthria communicate their emotions, how this may be encoded and thereby automatically extracted via prosodic and acoustic features, and how consistent a certain emotion is expressed among speakers and within the speaker him/herself (inter- and intra-speaker variability). All of these questions and probably more need to be addressed before we will be able to fully automate the process of identifying emotions in dysarthric speech and adding this information to the output of the AAC.

ACKNOWLEDGMENT

This research has been supported by a grant from the Saudi Ministry of Education, King Saud University, Saudi Arabia. The authors would like to thank Simon Judge (Senior Clinical Scientist, University of Sheffield, Rehabilitation and AT Group. Barnsley Assistive Technology Team) for his valuable comments, which have improved the design of the survey. The authors would also like to thank Communication Matters, the UK charity for augmentative and alternative communication for their help in distributing the survey and recruiting participants. A special thanks to all the participants for their time.

REFERENCES

- [1] M. Walshe and N. Miller, "Living with acquired dysarthria: the speaker's perspective," *Disability and rehabilitation*, vol. 33, no. 3, pp. 195–203, 2011.
- [2] J. R. Duffy, *Motor Speech Disorders-E-Book: Substrates, Differential Diagnosis, and Management*. Elsevier Health Sciences, 2013.
- [3] S. Steinberg, *Introduction to communication course book 1: The basics*. Juta and Company Ltd, 1995, vol. 1.
- [4] M. Fried-Oken, D. R. Beukelman, and K. Hux, "Current and future aac research considerations for adults with acquired cognitive and communication impairments," *Assistive Technology*, vol. 24, no. 1, pp. 56–66, 2012.
- [5] D. McNaughton and D. N. Bryen, "Aac technologies to enhance participation and access to meaningful societal roles for adolescents and adults with developmental disabilities who require aac," *Augmentative and Alternative Communication*, vol. 23, no. 3, pp. 217–229, 2007.
- [6] D. J. Higginbotham, H. Shane, S. Russell, and K. Caves, "Access to aac: Present, past, and future," *Augmentative and alternative communication*, vol. 23, no. 3, pp. 243–257, 2007.
- [7] M. S. Hawley, S. P. Cunningham, P. D. Green, P. Enderby, R. Palmer, S. Sehgal, and P. O'Neill, "A voice-input voice-output communication aid for people with severe speech impairment," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 21, no. 1, pp. 23–31, 2013.
- [8] R. Patel, "Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 6, pp. 1401–1415, 2003.
- [9] F. Miller and S. J. Bachrach, *Cerebral palsy: A complete guide for caregiving*. JHU Press, 2017.
- [10] D. R. Beukelman, S. Fager, L. Ball, and A. Dietz, "Aac for adults with acquired neurological conditions: A review," *Augmentative and alternative communication*, vol. 23, no. 3, pp. 230–242, 2007.
- [11] G. Pullin and S. Hennig, "17 ways to say yes: Toward nuanced tone of voice in aac and speech technology," *Augmentative and Alternative Communication*, vol. 31, no. 2, pp. 170–180, 2015.
- [12] A. Karpf, *The Human Voice: The Story of a Remarkable Talent*. Bloomsbury, 2007.
- [13] P. B. Dasgupta, "Detection and analysis of human emotions through voice and speech pattern processing," *arXiv preprint arXiv:1710.10198*, 2017.
- [14] D. A. Sauter, F. Eisner, A. J. Calder, and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion," *Quarterly Journal of Experimental Psychology*, vol. 63, no. 11, pp. 2251–2272, 2010.
- [15] R. Patel, "Prosodic control in severe dysarthria: Preserved ability to mark the question-statement contrast," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 5, p. 858, 2002.
- [16] R. Patel, "Phonatory control in adults with cerebral palsy and severe dysarthria," *Augmentative and Alternative Communication*, vol. 18, no. 1, pp. 2–10, 2002.
- [17] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-cultural elements in facial displays of emotion," *Science*, vol. 164, no. 3875, pp. 86–88, 1969.
- [18] R. P. Abelson and V. Sermat, "Multidimensional scaling of facial expressions," *Journal of experimental psychology*, vol. 63, no. 6, p. 546, 1962.
- [19] K. R. Scherer and G. Ceschi, "Lost luggage: a field study of emotion-antecedent appraisal," *Motivation and emotion*, vol. 21, no. 3, pp. 211–235, 1997.
- [20] I. S. Engberg, A. V. Hansen, O. Andersen, and P. Dalsgaard, "Design, recording and verification of a danish emotional speech database," in *Fifth European Conference on Speech Communication and Technology*, 1997.
- [21] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech," in *Ninth European Conference on Speech Communication and Technology*, 2005.
- [22] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The interface05 audio-visual emotion database," in *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*. IEEE, 2006, pp. 8–8.
- [23] P. Jackson and S. Haq, "Surrey audio-visual expressed emotion (savee) database," Apr 2011. [Online]. Available: www.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE/
- [24] S. Yacoub, S. Simske, X. Lin, and J. Burns, "Recognition of emotions in interactive voice response systems," in *Eighth European conference on speech communication and technology*, 2003.
- [25] M. Luggner and B. Yang, "The relevance of voice quality features in speaker independent emotion recognition," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4. IEEE, 2007, pp. IV–17.