What does it mean to communicate (not) emotionally?

Jan-Oliver Wülfing
Fraunhofer Centre Birlinghoven IZB
Schloß Birlinghoven
D-53754 Sankt Augustin
`jan-oliver.wuelfing`
`@izb.fraunhofer.de`

Lisa Hoffmann
Fraunhofer Institute for
Applied Information Technology FIT
Schloß Birlinghoven
D-53754 Sankt Augustin
`lisa.hoffmann@fit.fraunhofer.de`

Abstract

Communication is an essential part of our life. Though, not only communication is the key – it is all about emotional (prosodic) communication. Due to empirical research, people, who are augmentative communicators and speak with a voice output communication aid, want to express their emotions in the same way as everybody else – it is one of their deepest interests (Portnuff, 2006; Hoffmann and Wülfing, 2010). So far, current devices lack the opportunity of emotional utterances. This circumstance leads not only to a huge usability deficit, but furthermore, it is an obstacle to develop emotional competence and to behave as well as regulate one´s emotion adequately (Blackstone and Wilkins, 2009). This article aims to increase the sensitivity for the importance of emotional communication. Furthermore, it tries to give first suggestions for implementing an usable device that supports users with a voice output communication aid to express emotional utterances. This could be done by using phrase-generation, as mentiond by Vanderheyden and Pennington (1998).

1 Introduction

One can hardly imagine how it would be to communicate without an emotional output. If we talk to each other, every conversation is influenced by our emotions. Sometimes we want to hide our feelings; however, more often we want to send an underlying message with our prosodic utterance. For example, imagine someone goes for a walk with their best friend, the sun is shining and the birds are chirping – this kind of situation and the feeling it evokes will probably be reflected by their prosodic utterance: "Dude, it`s awesome weather today!" Or imagine furthermore, the same person gets into a fight with this friend while walking in the sun; it certainly must not be pointed out that the emotion and therefore the reaction would differ. Though, communication and emotion seem inevitably associated, it is still not possible for people with complex communication needs to talk emotionally, since current voice output communication aids (VOCA) do not offer prosodic speech output. A circumstance that leads to several drawbacks; starting with disadvantages in social and academic arrangements.

1.1 What is a prosodic utterance?

Prosodic utterances are the key for daily communication processes. They depend on the emotion, i.e. it is reflected by an utterance through the pitch, the rhythm, and the volume of the voice. These differences are called prosody and, hence, it is possible to express very different types of emotions. Prosodic utterances influence the behaviour of the listener (Strom, 1998). The one and the same utterance can differ in their meaning. A good example of this is irony or subtle utterances. They may or may not be serious depending also on their prosodic features. Either way, the listener is going to react and this reaction relies not only on the accurate comprehension but also on the prosody of the speaker´s intended utterance (Battachi et al., 1997). In the case of augmentative communicators and their use of a

VOCA, these aspects are not fully fulfilled. Instead of underlining their utterances with one or more prosodic features in order to convey their emotions to the listener, they must transcribe their current emotion in text input of the VOCA. In turn, the VOCA renders this in a monotonic, mostly synthetic voice output, which leads to two objections:

Firstly, the listener misinterprets the augmented communicator´s utterance, which may have dramatically effects for a follow-up conversation. Secondly, the listener becomes just bored and the conversation does not last long. In both ways, the augmented communicator´s situation becomes worse, since it influences their social environments and, thus, themselves (Balandin, 2005).

2 Emotional competence as a pre-condition for social participation

Emotions are seen to be an essential factor of social communication. To be a part of social relationships, it is necessary to develop emotional competence in some kind. Janke (2008) postulates emotional competence as the ability to express and regulate the own emotions and, furthermore, it describes the ability to understand emotions – the own one´s and other one´s. However, without the possibility to talk emotionally, it is hard to develop emotional competence. Research, in fact, has shown that users of a VOCA often have deficits in this area which leads to difficulties in forming relationships and the adequate emotion regulation is influenced as well (Brinton and Fujiki, 2009; Blackstone and Wilkins, 2009). Furthermore, there is a significant correlation between children´s emotional knowledge and the degree of peer popularity (Janke, 2008). Brinton and Fujiki (2009) even suggest that emotional competence has impact on academic learning. Due to the fact that the development of emotional competence is learned in day-to-day interactions with other people and that emotional utterances are a necessary pre-condition for exactly these interactions, it becomes clear that the development of a VOCA which does support prosodic communication would be an important factor for increasing the user´s Quality of Life. Among other things, it includes warm interpersonal relationships and academic achievements.

2.1 Development of social behaviour

Recent psychological theories focus on mutual-information processing systems for explaining social behaviour. Strack and Deutsch (2004) postulate two determinants to guide one´s decision-making: the reflective and the impulsive system. Though, both systems are interacting, they are different in their characteristics and functioning. The reflecting system is rather rational; it includes knowledge about facts and (social) values on which it bases its decision. The impulsive system, on the other hand, is lacking rational reasoning. It is rather impulsive, quick, and affected by motivational factors. Whereas the impulsive system is part of every decision making process, the reflective system is not. As, for instance, it needs more cognitive resources while operating and is easily disturbed.

When transferring the model onto emotional processing and electronic communication aids, it appears that alternative communicators are challenged in both, the impulsive and the reflective system. By definition, emotions are impulsive, quick, and the decisions based on them are often lacking rational reason. Thus, most of the time, emotional behaviour is driven by the impulsive system. Due to slow input-rates, users of a VOCA, indeed, are not able to communicate their emotions quickly and impulsively. They have to rely on the reflecting system. In some cases one might argue that this is the better opportunity as impulsive emotional utterances are regretted at times. On that account children learn that in some situations it is important to not follow their (inappropriate) impulsive behaviour while growing up (Blackstone and Wilkins, 2009). But due to the fact that alternative communicators are often disadvantaged in developing appropriate emotional behaviour (Brinton and Fujiki, 2009; Blackstone and Wilkins, 2009), it is also difficult to provide an adequate basis of knowledge for the reflective system. Thus, in particular for children, is important to communicate impulsively as it also strengthens the ability to make rational choices. Taking Strack´ and Deutsch´s theory (2004) into account it becomes clear that for the purpose of impulsive reactions an intended prosodic VOCA requires a possibility for a fast input-rate.
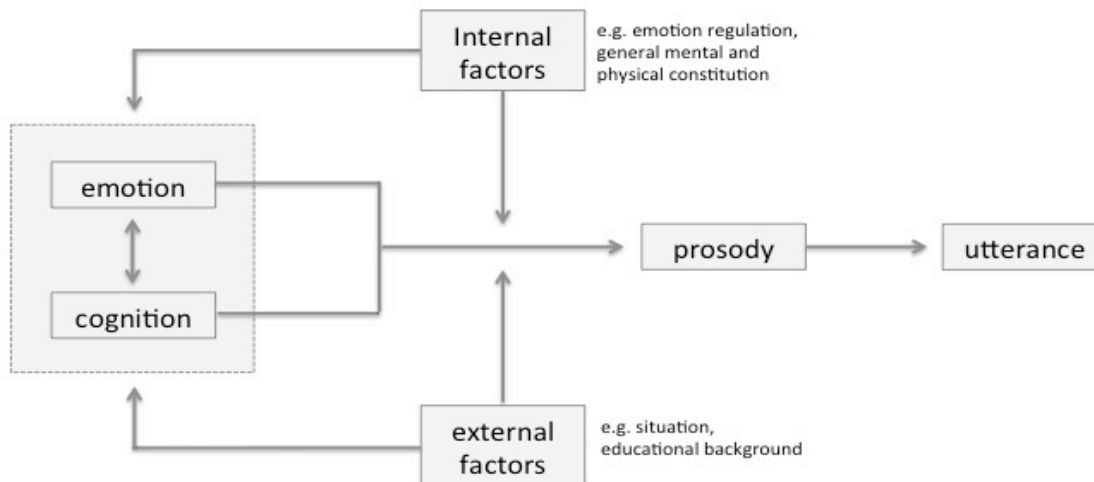
Fig. 1: The origin of prosodic utterances which reflects the emotional state. Internal and external factors have both influence on the emotion and cognition development as well as on the utterance itself.

3   Usability deficits

Empirical research in Usability Engineering shows that users of a VOCA claim for emotional communication (Hoffmann and Wülfing, 2010). Though, they typically honour the prospects given by the devices, they still describe the missing opportunity to talk emotionally (Portnuff, 2006; Hoffmann and Wülfing, 2010). If we take Maslow´s (1970) 'Hierarchy of Needs' into account, it is indeed not surprising that people, who have complex communication needs, want to talk in a very normal way. Maslow´s purpose is to show that every human being has specific needs. These needs are ordered hierarchically. The lowest frame depicts physiological needs like nutrition, sex, and the activity of the autonomous nervous system which are mostly involuntary (e.g. breathing). Then, the next layer symbolises all the aspects of safety. If these needs are fulfilled, it comes to friendship and love needs, those include emotional talk, social relationships, and emotional competence. His hierarchy underlines the assumption that a communication without prosody is acceptable but not satisfying, since emotional talk is an essential factor for being part of social relationships. Here, we go one step forward, because if we say that a product must be usable as

described in ISO 9241-11 (1998), it must also be satisfying. The ISO standard defines Usability as: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." The effectiveness means the accuracy in achieving the specific goal. Efficiency is time and effort the user needs for the achievement. And, satisfaction is the positive attitude towards the system. The three aspects are arranged hierarchically as shown in figure 2. Taking this definition into account and extending it on electronic communication aids, it becomes clear that the missing possibility for emotional communication can not only be seen as a satisfactory-failure but also as an effective-failure: the user is not able to achieve their goal accurately. Imagine, for instance, a human being cannot move because they have a significant impairment. The person is sitting in a wheelchair in one corner of a room and has a VOCA mounted on the wheelchair. Via the device the user asks the people in the room to move their chair, however, no one listens and helps due to the monotonous voice. In this scenario the VOCA fails and the user is not able to achieve their goal. The device, therefore, is not effective. But would this happen if the VOCA would have a prosodic voice output? We believe that this would

96

not happen. With a prosodic voice it is possible to get attention more easily. In a normal situation, if we could use our voice, we would rather become angry more and more if no one listens to us. Our cadence would unveil our emotion (Scherer and Banse, 1991) and, hence, would underline our intention. For augmented communicators, this would be an important step forward to not only communicate the semantic content of their utterances, but also the emotion underlying those. The actual relevance of the made-up scenario above is also supported by VOCA user Collin Portnuff. He claimed in a speech, given in 2006, a device capable of shouting to gain attention in certain situations (Portnuff, 2006).

Fig. 2: Hierarchical steps of Usability based on ISO 9241-11 (1998)

The elimination of the uability deficit described above would also strengthen the user´s position in a communication and most likely the duration of a communication. We all love more to talk to people whose voice contains prosodic features than to face anyone whose voice is rather monotonous. Due to the fact that human beings are empathic, we like to have an impression of the opposite´s emotion. And how do we get such an impression? Through the mimic, gesture, and the tone of voice. So, an emotional device is a Win-Win situation: The listener can respond adequately and the speaker can express his intention through a verbal prosodic utterance. It becomes clear that even from the listener´s point of view it is easier and more comfortable to have a conversation with an augmented communicator who has a prosodic VOCA, since it would fulfil at least some constraints of a fluent conversation (Todman and Alm, 2003). Or as Collin Portnuff (2006) puts it: "And when you help someone communicate, you

are not just helping that person, but all the people with whom he or she interacts."

4   Often heard criticism

Often heard criticism about a prosodic speech output of a VOCA typically embraces the following three aspects:

a. Emotional conversation would increase the input rate,

b. emotional utterances are not possible without a synthetic voice that supports emotional output and

c. emotions and emotional conversations are too complex to work on a VOCA.

Even though these aspects are not to be dismissed, we assume that it is possible to find proper solutions for each of them.

a. Increasing input rate: The speech generic device of a VOCA with a specific prosodic tool should enable the possibility of an emotional conversation without increasing the input time. Instead, emotional utterances will extend the duration of a conversation. So far augmentative communicators use their VOCA often to communicate common needs as "I´m hungry" or to answer simple yes-/no-questions (Blackstone and Wilkins, 2009). Lasting conversations as for example talking about a film watched just before at a cinema and the experienced emotions while doing so are rather rare. In the latter kind of conversations emotional utterances are very important, since a lack of them would shorten a conversation dramatically. Also one can guess that the user´s need to have an emotional equipped VOCA is higher than the expense of one more additional keystroke. Nevertheless it is important to keep any additional effort to the lowest to truly fulfil the user´s requirements. However, the missing lasting conversations are to be seen as a gap in the Quality of Life. In order to bridge this gap along with the gaps mentioned above (e.g. the missing emotional competence and Usability deficits) the development of a VOCA, that does support prosodic communication, seems inevitable.

b. Synthetic voice: Starting with a prototype of VOCA, the emotional utterances can easily be pronounced naturally instead of using synthesized

97

speech, since emotional speech is rather limited in current VOCAs (Higginbotham, 2010). Nonetheless, there has been some notable efforts in recent years (e.g. CereProc Ltd.[1]). They make use of what they call 'Emotional Continum'. It is possible to simulate a wide range of underlying emotions of the voice. Or, if we look at the work of the World Wide Web Consortium, they currently work on the EmotionML (2011) which should facilitate synthetic voices to become annotated with emotional tags. So, it is to assume that in the near future emotional synthetic utterances will be possible. In the meantime, however, a natural recorded voice of an actor is an acceptable solution as current existing synthetic voices are often experienced as alien (Hoffmann and Wülfing, 2010). Furthermore, user experience shows some people find it difficult to listen to the same intonations given by the devices while the meaning of the words change (Portnuff, 2006), a problem that can be solved by using natural voice output.

c. Complexity: In order to limit the complexity of such a formation, it is necessary to start with isolating a reasonable amount of emotions. It is also important to not use indifferent emotions, as for example cold anger and panic fear, as people seem to have problems distinguishing them (Banse and Scherer, 1996) which may lead to frustration in conversations later on. The three emotions happiness, anger, and sadness are quite different in their prosody. Thus, a listener can recognise them very well, as Burkhardt (2001) mentioned. They belong to the so-called basic emotions as well (Ekman, 1999). Therefore, it seems to be a reasonable choice to choose them for a start-up project. Another possibility to lower the complexity is to attach the emotions to certain situations; this restricts the context of utterances.

5 Initial approaches

The proposed project aims at a spontaneous emotional communication in context-specific situations for VOCA users. This includes the idea of identifying a user-specific and context-specific vocabulary based on phrase-generation. As shown above there are certain requirements for an authentic emotional communication: Any emotional utterance consists of its semantic

content, its prosodic characteristic, and a certain degree of impulsiveness. The presented initial approaches keep these requirements in mind as well as the beforehand mentioned criticism. It is important to note that these initial approaches are first propositions which are based on current knowledge. Possibly some adjustments need to be made in the development of the prosodic VOCA in order to keep the device truly usable.

Yet, based on current experiences we propose a system where users firstly select their emotion and secondly compose the prosodic utterance, since the emotion does typically not change after each utterance. Thus, the user does not need to change the prosodic filter option every time. In this way, the input-rate will not enhance unnecessarily as often criticised. To render the possibility of impulsive, spontaneous, and agile communication, the prosodic VOCA needs to support the user with a sample of potential utterances fitting the user´s emotion in the specific situation. Therefore, the development of a phrased-based vocabulary is necessary. It is important to guarantee the validity of the possible utterances given by the device as otherwise they will not be perceived authentic. That is why the potential samples of utterances must not be chosen at random. Instead they should be based on empirical settings. This will be done by investigating specific contexts and identifying emotional phrases given by a specific subpopulation. The probability enhances that the device offers the user an utterance which he or she actually needs in the specific situation by using empirical based methods. The established vocabulary should be user-specific by all means. There are age-based differences in the way people speak and express their emotions. A user-specific vocabulary needs to keep these age-differences in mind for identifying phrases that fit the phraseology of specific users. Therefore, the sample of potential utterances should also be based on colloquial speech. This is a good example in order to enrich a conversation more lively and ongoing. In particular, colloquial speech enables the augmentative communicators to be perceived more authentic by their social environment. It also supports the user in developing emotional competence using impulsive speech. In turn the environment´s feedback increases the adequate

---

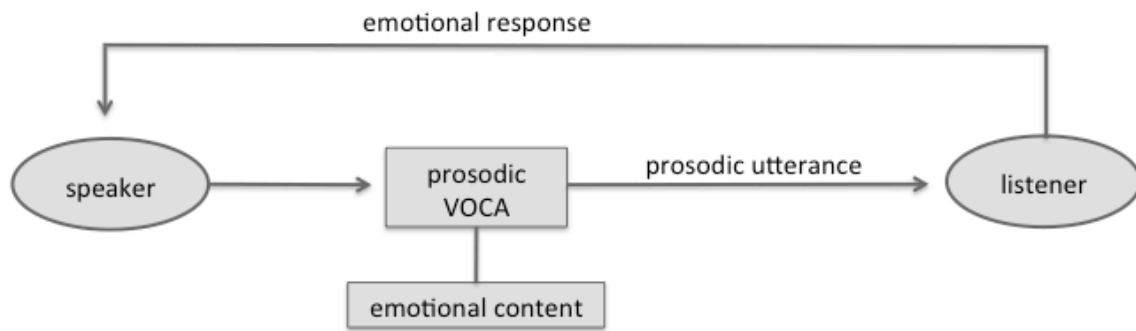[1] www.cereproc.com (accessed 06/24/2011)

Fig. 3: Model of emotional communication using a prosodic Voice Output Communication Aid which allows the listener an appropriate emotional response to the speaker`s (augmented communicator) intented message.

handling of emotions in specific situations (s. Fig. 3). In addition to the possibility of using the utterances given by the device the user still needs the alternative to form contents individually. A prosodic VOCA does not aim at depriving the users of the power to speak independently. It rather serves as an adequate support in order to enhance the promptness of a statement which is an important requirement for an authentic conversation. As already mentioned, natural recorded voices could be used for the potential utterances as an alternative to synthetic speech output. However, it is important to note that the utterances must resemble the appropriate prosody. Hence, the pitch, the rhythm, and the volume of the voice need to fit the content of the utterance. Sorrow e.g. should be presented rather quiet while enragement requires a higher sound level.

All things considered, it becomes clear that the development of a prosodic VOCA goes along with some important requirements that reflect the user´s needs. In order to develop a device, which supports these needs, the Usability has to be seen as an essential factor. Thus, it is important to include the user throughout the whole process of developing. A constraint for designing usable devices that fit the Usability definition is the ISO standard 9241-11 (1998). This kind of Usability-Setting include interviews with users and when indicated a monitoring with a 'Thinking aloud-'method, which helps to formalise specific usability problems with the specific system in that specific context.

Summarised, the requirements for a prosodic VOCA should be based on:

- phrase-generation
- specific contexts
- appropriate emotions
- user-specific phraseology
- adequate prosody
- usability standards

To validate these and additional assumptions for a specific sample of users, they must be confirmed in an emprical setting. A frist step was done by showing the importance of emotional utterances for augmented communicators (Hoffmann and Wülfing, 2010).

6 Conclusion

Emotional communication is an essential part of everyday life - this is true for people with and without disabilities. Lacking the opportunity of talking emotional means to miss out many aspects of a fulfilling life, since emotional output has an enormous impact on social relationships, the developing of emotional competence, and even on academic achievements. Furthermore, this circumstance leads to a huge usability deficit. Augmented communicators´ own expressiveness of emotions by gesture and mimic means is limited and prosodic communication is not possible, yet. Precisely because, prosodic VOCA is a real

innovation. It gives people with complex communications needs the opportunity to express themselves emotionally. It would encourage their participation in social life and, thus, also their Quality of Life. Still, there is a lot of criticism involving this topic, however, with adequate methods and ideas it does seem possible that some day in the future users of a VOCA will be able to communicate emotionally.

References

Susan Balandin (2005). Ageing and AAC: Start early, age well! In Jens Boenisch and Karin Otto (eds.), Leben im Dialog – Unterstützte Kommunikation über die gesamte Lebensspanne (466-478). Von Loeper, Karlsruhe.

Marco W. Battachi, Thomas Suslow, and Margherita Renna (1997). Emotion & Sprache – Zur Definition der Emotion und ihren Beziehungen zu kognitiven Prozessen, dem Gedächtnis und der Sprache. Peter Lang, Frankfurt a.M.

Rainer Banse and Klaus R. Scherer (1996). Acoustic profiles in vocal emotion expression. In: Journal of Personality and Social Psychology, 70(3), 614-636.

Sarah W. Blackstone and David P. Wilkins (2009). Exploring the Importance of Emotional Competence in Children With Complex Communication Needs. In: Perspectives on Augmentative and Alternative Communication, 18, 78-87.

Felix Burkhardt (2001). Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren. In: Reihe Berichte aus der Kommunikationstechnik. Shaker, Aachen.

Bonnie Brinton and Martin Fujiki (2009). Meet me more than half way: Emotional competence in conversation using AAC. In: Perspectives on Augmentative and Alternative Communication, 18, 73-77.

Paul Ekman (1999). Basic Emotions. In: Tim Dalgleish and Mick Power (Eds.), Handbook of Cognition and Emotion. John Wiley & Sons, Sussex, U.K.

Emotion Markup Language 1.0 (2011) W3C working draft. http://www.w3.org/TR/2011/WD-emotionml-20110407/ (accessed 06/24/2011).

Jeffrey Higginbotham (2010). Humanizing Vox Artificialis: The Role of Speech Synthesis in Augmentative and Alternative Communication. In: John Mullennix and Steven Stern (Eds.), Computer Synthesized Speech Technologies – Tools for Aiding Impairment (50-70). IGI Global, Hershey, PA.

Lisa Hoffmann and Jan-O. Wülfing (2010). Usability of Electronic Communication Aids in the Light of Daily Use. In: Proceedings of the 14th Biennial Conference of the International Society for Augmentative and Alternative Communication (259). Barcelona, Spain

ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.

Bettina Janke (2008). Emotionswissen und Sozialkompetenz von Kindern im Alter von drei bis zehn Jahren. In: Empirische Pädagogik, 22(2), 127-144.

Abraham H. Maslow (1970). Motivation and Personality. Harper & Row, New York, NY.

Colin Portnuff (2006). AAC: A user´s perspective. Webcast available as part of the AAC-RERC Webcast Series. http://aac-rerc.psu.edu/index-8121.php.html (accessed 03/30/2011).

Klaus R. Scherer, Rainer Banse, Harald G. Wallbott, and Thomas Goldbeck (1991). Vocal Cues in Emotion Encoding and Decoding. In: Motivation and Emotion, 15(2), 123-148.

Fritz Strack and Roland Deutsch (2004). Reflective and Impulsive Determinants of Social Behavior. In: Personality and Social Psychology Review, 8(3), 220-247.

Volker F. Strom (1998). Automatische Erkennung von Satzmodus, Akzentuierung und Phrasengrenzen. PhD thesis, University of Bonn.

John Todman and Norman Alm (2003). Modelling conversational pragmatics in communications aids. In: Journal of Pragmatics, 35, 523-538.

Peter B. Vanderheyden and Christopher A. Pennigton (1998). An Augmentative Communication Interface Based on Conversational Schemata. In: Vibhu O. Mittal, Holly A. Yanco, John Aronis, Richard C. Simpson and Richard Simpson (Eds.): Assistive Technology and AI, LNAI 1458, 109-125.