Towards Monitoring of Children's Speech - A Case Study

Andreas Maier
Abteilung für Phoniatrie und Pädaudiologie, Universität Erlangen-Nürnberg
Martensstr. 3
Erlangen, Germany
andreas.maier@cs.fau.de

Maria Schuster
Abteilung für Phoniatrie und Pädaudiologie, Universität Erlangen-Nürnberg
Bohlenplatz 21
Erlangen, Germany
maria.schuster@uk-erlangen.de

Elmar Nöth
Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg
Martensstr. 3
Erlangen, Germany
elmar.noeth@cs.fau.de

ABSTRACT

For speech disorders, state-of-the-art assessment is performed mainly perceptive. This causes a lot of effort. Speech processing techniques are applicable to perform an automatic and observer-independent evaluation of the speech intelligibility. High agreement between human raters and the automatic evaluation method were found. In this study we investigate, whether the automatic evaluation system is also suitable to perform speech intelligibility follow-up. Speech data of six children with cleft lip and palate (CLP) were investigated with one year in between. The results show that the annual improvement in these children was in the same range as in the the age-matched control group on average. The automatic evaluation is in agreement with perceptual evaluation.

1. INTRODUCTION

Communication has become more and more important for our daily life, due to the change from the industrial age to the information age. About 87.5 % of the inhabitants of urban areas require communication for their daily work. Communication disorders cause a major effect on the economy. The cost of care as well as the degradation of the employment opportunities for people with communication disorders cause a loss of $154 billion to $186 billion per year to the economy of the United States of America alone. This equals to 2.5 % to 3.0 % of the Gross National Product of the US. These facts indicate that communication disorders are a major challenge in the 21st century [11]. The use of automatic speech processing techniques will contribute to ameliorate diagnostics and therewith will identify the best therapy methods. Hence, it increases the employment opportunities for people who suffer from speech disorders.

CLP might cause communication disorders. In this work we investigate whether automatic speech recognition is suitable for the speech follow-up of children with CLP as a first step towards better rehabilitation of communication disorders.

2. METHODS

The "Program for the Evaluation of All Kinds of Speech disorders" (PEAKS, [8]) was applied to record and analyze the data. It is an online speech recording and evaluation system which is available by internet (http://peaks.informatik.uni-erlangen.de).

In a first acoustic analysis, the speech recognizer converts speech into a sequence of feature vectors which consist of 12 Mel-Frequency Cepstrum Coefficients (MFCC). The first coefficient is replaced with the energy of the signal. Additionally 12 delta coefficients are computed over a context of 2 time frames to the left and the right side (50 ms in total).

The recognition is performed with semi-continuous Hidden Markov Models. The codebook contains 500 full covariance Gaussian densities which are shared by all HMM states. The elementary recognition units are polyphones [14], a generalization of triphones. Polyphones use phones in a context as large as possible which can still statistically be modeled well, i.e., the context appears more often than 50 times in the training data. The HMMs for the polyphones have three to four states per phone.

We used a unigram language model to weigh the outcome of each word model. It was trained with the reference words of the tests. Only the frequency of occurrence for each word in the used text was known to the recognizer, since for our purpose it was necessary to put more weight on the recognition of acoustic features.

The speech recognition system had been trained with acoustic information from spontaneous dialogues of the VERBMOBIL project [17] and normal children's speech. The speech data of non-pathologic children's voices (30 female and 23 male) were recorded at two local schools (age 10 to 14) and consisted of read texts. The training population of the VERBMOBIL project consisted of normal adult speakers from all over Germany and thus covered all regions of dialect. All speakers were asked to speak "standard" German. 90 % of the training population (47 female and 85 male) were younger than 40 years. During training an evaluation set was used that only contained children's speech. The adults' data was adapted by vocal tract length normalization as proposed in [16].

MLLR adaptation [3, 6] with the patients' data lead to further improvement of the speech recognition system.

From the recognized word chain and the reference a percentage of correctly uttered words — the word recognition

Table 1: Overview on the distribution of the WR in the control group

age	controls		
	#	WR	diff. to prev. year
six	61	54.2 ± 7.3	-
seven	188	59.1 ± 9.0	4.9
eight	155	63.2 ± 9.1	4.1
nine	127	65.8 ± 8.5	2.6
ten	94	68.0 ± 8.7	2.2

rate (WR) — is computed. As previously shown this number corresponds with the speech intelligibility [15], i.e., since the recognizer and the setup is kept constant the only varying factor remains the child's speech that is recorded. A high correlation (r=0.89) between the recognition accuracy and the speech intelligibility of children with CLP was reported [5].

3. DATA

CLP is the most common malformation of the head with incomplete closure of the cranial vocal tract [18, 9, 10, 13, 12]. Speech disorders can still be present after reconstructive surgical treatment. The characteristics of speech disorders are mainly a combination of different articulatory features, e.g. enhanced nasal air emissions that lead to altered nasality, a shift in localization of articulation, and a modified articulatory tension, e.g. weakening of the plosives [4]. They affect not only the intelligibility but therewith the social competence and emotional development of a child.

Acoustic files were recorded from six children with CLP. One of them had an isolated cleft lip while the other children had unilateral cleft lip and palate [1]. The examination was included in the regular out-patient examination of all children with CLP. Informed consent had been obtained by all parents prior to the examination. All children were native German speakers using a local dialect.

The children were asked to name pictures that were shown according to the PLAKSS test [2]. This German test consists of 99 words. It includes all possible phonemes of the German language in different positions (beginning, center and end of a word). Figure 1 shows an example slide for the phoneme /r/ with the German target words T**r**ecker, Zit**r**one, Jäge**r** (tractor, lemon, hunter).

In order to compare the patients' speech to normal children's speech a large control group of 625 children was gathered in four German major cities, which cover most regions of dialect [7]. Table 1 shows the age-dependent normative values which were determined with this control group.

4. RESULTS

Table 2 shows the children of the patient group. The children were recorded twice with one year in between. Half of the children gain 3 to 4 percent points which corresponds to the control population's gain (cf. Table 1). However, one child (ID 12163) shows even degradation. Perceptual evaluation confirmed this finding. The child with no progress already had a low WR at the first recording (34.3 %) which is far below the age-matched average (63.2 ± 9.1 %). Two of the children (IDs 15017 and 17661) improved above the age-matched average. Again, this finding could be perceptually confirmed. Especially, ID 15017 showed a lot of improvement. On average the CLP children gained 3.7 percent points which is in the same range as the children in the control group. However, the variance is very high, especially in the second recording, since some children improve and some don't.

Table 2: The children in the dataset were recorded a second time after one year. Hence, their progress within this year can be compared

ID	age			WR		
	1st	2nd	diff.	1st	2nd	diff.
21402	6.0	6.9	0.9	45.5	48.5	3.0
19646	6.7	7.8	1.1	31.3	34.3	3.0
17661	6.9	8.0	1.1	44.4	52.5	8.1
15552	7.7	8.7	1.0	38.4	34.3	4.0
15017	7.9	8.9	1.0	51.5	69.7	18.2
12163	9.3	10.1	0.8	34.3	28.3	-6.1

5. DISCUSSION

In this paper the application of ASR is presented to the follow-up of speech development in CLP children. A problem for all such applications is that children's speech develops. Hence, not only the WR but also the children's speech in general is age-dependent. As in language development tests age-matched standard values have to be obtained. Then, an evaluation of the speech data can be performed with respect to the age of the child.

Furthermore, human speech is complex. Secondary disorders such as attention deficit disorders may also influence the recognition accuracy. Hence, they have to be taken into account as well when the figures are interpreted.

In general, the follow-up within the regular out-patient examination is sometimes problematic, since children whose speech disorders improved do not return to the regular examination. Only the children with disorders return regularly. Hence, it is difficult to gather data continuously of a large group of patients.

With the age-dependent WR values as presented in Table 1 we are able to interpret whether the improvement of the child was similar to the children in the same age. If improvements were made during the therapy the increase of the WR should be above the average increase in age-matched children.

The average improvement of the CLP children in one year was in the same range as in the control group. However, the variance was high because some children increase their intelligibility while others don't. Both children that improved above-average had speech therapy during the year. Further-

Figure 1: Pictograms of the PLAKSS test for the phoneme /r/ with the German target words Trecker, Zitrone, Jäger (tractor, lemon, hunter).

more, the child with ID 15017 had a attention deficit disorder which was successfully treated. The speech therapy was also successful. Only in one severe case degradation occurred. The child had a mental retardation (IQ 69). Unfortunately, speech therapy could not improve the child's speech intelligibility yet.

The investigated group is too small to find significant differences within one year, however, all findings were perceptually confirmed. Hence, the method is suitable for documentation and follow-up of children's speech disorders.

6. SUMMARY AND OUTLOOK

In this paper we presented a case study which involved six children with cleft lip and palate in comparison to a large age-matched control group. It was shown, that the children improve their recognition rates on average at the same rate as children of a control group. The changes in the individual children differed a lot. All automatically evaluated results could be confirmed perceptually. The method is easy to apply and allows an objective quantification of the speech intelligibility.

A long term investigation in the future will bring more detailed results on the development of speech disorders in CLP children.

In this study only children in school age could be investigated. The gathering of more control data will allow the investigation of preschool children and high school children in the future. Therefore, age-depended WR values have to be collected and standardized.

7. REFERENCES

[1] B. Eppley, J. van Aalst, A. Robey, R. Havlik, and M. Sadove. The spectrum of orofacial clefting. *Plastic and Reconstructive Surgery*, 115(7):101–114, 2005.

[2] A. Fox. PLAKSS – Psycholinguistische Analyse kindlicher Sprechstörungen. Swets & Zeitlinger, Frankfurt a.M., Germany, now available from Harcourt Test Services GmbH, Germany, 2002.

[3] M. Gales, D. Pye, and P. Woodland. Variance compensation within the MLLR framework for robust speech recognition and speaker adaptation. In *Proceedings of the International Conference on Speech Communication and Technology (Interspeech)*, volume 3, pages 1832–1835, Philadelphia, USA, 1996. ISCA.

[4] A. Harding and P. Grunwell. Active versus passive cleft-type speech characteristics. *Int J Lang Commun Disord*, 33(3):329–52, 1998.

[5] A. Maier, C. Hacker, E. Nöth, E. Nkenke, T. Haderlein, F. Rosanowski, and M. Schuster. Intelligibility of children with cleft lip and palate: Evaluation by speech recognition techniques. In *Proc. International Conf. on Pattern Recognition (ICPR)*, volume 4, pages 274–277, Hong Kong, China, 2006.

[6] A. Maier, T. Haderlein, and E. Nöth. Environmental Adaptation with a Small Data Set of the Target Domain. In P. Sojka, I. Kopeček, and K. Pala, editors, *9th International Conf. on Text, Speech and Dialogue (TSD)*, volume 4188 of *Lecture Notes in Artificial Intelligence*, pages 431–437, Berlin, Heidelberg, New York, 2006. Springer.

[7] A. Maier, T. Haderlein, E. Nöth, and M. Schuster. PEAKS: Ein Client-Server-Internetportal zur Berwertung der Aussprache. In *Telemed 2008, Proceedings*, pages 104–107, Heidelberg, Germany, 2008. Akademische Verlagsgesellschaft, Aka GmbH.

[8] A. Maier, E. Nöth, A. Batliner, E. Nkenke, and M. Schuster. Fully Automatic Assessment of Speech of

Children with Cleft Lip and Palate. *Informatica*, 30(4):477–482, 2006.

[9] T. Millard and L. Richman. Different cleft conditions, facial appearance, and speech: relationship to psychological variables. *Cleft Palate Craniofac J*, 38:68–75, 2001.

[10] F. Rosanowski and U. Eysholdt. Phoniatric aspects in cleft lip patients. *Facial Plast Surg*, 18(3):197–203, 2002.

[11] R. Ruben. Redefining the survival of the fittest: communication disorders in the 21st century. *Laryngoscope*, 110(2):2410245, 2000.

[12] R. Schönweiler, J. Lisson, B. Schönweiler, A. Eckardt, M. Ptok, J. Trankmann, and J. Hausamen. A retrospective study of hearing, speech and language function in children with clefts following palatoplasty and veloplasty procedures at 18-24 months of age. *Int J Pediatr Otorhinolaryngol*, 50(3):205–217, 1999.

[13] R. Schönweiler and B. Schönweiler. Hörvermögen und Sprachleistungen bei 417 Kindern mit Spaltfehlbildungen. *HNO*, 42(11):691–696, 1994.

[14] E. Schukat-Talamazzini, H. Niemann, W. Eckert, T. Kuhn, and S. Rieck. Automatic Speech Recognition without Phonemes. In *Proceedings European Conference on Speech Communication and Technology (Eurospeech)*, pages 129–132, Berlin, Germany, 1993.

[15] M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, and E. Nöth. Evaluation of Speech Intelligibility for Children with Cleft Lip and Palate by Automatic Speech Recognition. *Int J Pediatr Otorhinolaryngol*, 70:1741–1747, 2006.

[16] G. Stemmer, C. Hacker, S. Steidl, and E. Nöth. Acoustic Normalization of Children's Speech. In *Proc. European Conf. on Speech Communication and Technology*, volume 2, pages 1313–1316, Geneva, Switzerland, 2003.

[17] W. Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer, New York, Berlin, 2000.

[18] N. Wantia and G. Rettinger. The current understanding of cleft lip malformations. *Facial Plast Surg*, 18(3):147–153, 2002.