ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

OBJECTIVE LONG-TERM ASSESSMENT OF SPEECH QUALITY CHANGES IN PRE-LINGUAL COCHLEAR IMPLANT CHILDREN

Hamid Sheikhzadeh and Rassoul Amirfattahi

Dept. of Elect. Eng., Amirkabir Univ. of Tech. (Tehran Polytechnic), Tehran, Hafez Ave., Iran 15914

Email: hsheikh@cic.aku.ac.ir

ABSTRACT

In this research, we have studied the static features of three main Farsi vowels, uttered by four pre-lingual cochlear implant (CI) children. Speech samples have been studied before the CI operation and every three months, up to nine months post-operation. To assess the effects of the auditory feedback (AF) on the speech quality, patients have spoken with their CI devices in both on and off positions. Quantitative results show that 1) after nine months post-operation, almost all of the static features converge towards their normal values, and 2) the dependency of the static features on the AF decreases with time, which implies that the speech production motor patterns of the patients have been trained in time.

1. INTRODUCTION

The evaluation of time-developments of speech produced by CI patients has been the focus of many recent researches [1-5]. The issue is of particular importance in the case of CI children with pre-lingual deafness, since (unlike post-lingual children) drastic changes occur in their speech qualities. Thus the scope of this study is limited to pre-lingual group of CI patients. Through careful speech quality assessments, researchers can address a few important research topics including the following important issues:

First, as a patient's speech quality is an indication of how well he/she hears, one can ultimately use the speech quality evaluations to adjust the patient's CI device. This approach is especially helpful in the case of young children who cannot give adequate feedback about what they hear, but are able to speak.

Second, the study of role of the AF on human speech production mechanism has been an active research topic for many years. Through careful examination of speech produced by CI patients in device on and off situations, one can obtain more objective answers to the problem. Of course the issue is far too complicated to be fully addressed by a few studies. As this research shows, the role of AF varies by time, and furthermore, there are so many static and dynamic features of speech that are affected by the AF in different, and sometimes-conflicting ways.

Third, different coding strategies have been employed in CI. Systems. One can objectively evaluate and compare various coding strategies by careful speech quality assessment for patients using different strategies.

Most of the researchers who have studied the speech quality of the pre-lingual CI children, have chosen to compare the speech quality in on/off position of the CI device, at a certain time after operation. The short-term nature of their approaches, and their difference in time periods of the tests, have led to sometimes-contradictory results [1-4]. In this paper, we report the results of a long-term study, as well as a device on/off study that is repeated every three months. By considering the time variation of the results, some of the previous inconsistencies in the reported conclusions can be clarified.

2. RESEARCH METHOD

The group under study consisted of four Farsi-speaking children (all females, 7-13 years old and called AZ, MA, NA, and EL in this paper) with pre-lingual deafness, all using NUCLEUS-CI24M devices [6] with SPEAK coding strategy. The speech quality assessment is limited to three tense Farsi vowels of /a/, /i/ and /u/ that exist in almost all languages. Five meaningful sentences uttered by the patients were recorded, before the operation, and three, six, and nine months post-operation. At each session, patients read the sentences once in prosthesis-on (PN) condition and then after 30 minutes of stay in a prosthesis-off (PF) condition. To be able to conduct a more objective study (rather than qualitative and subjective evaluations), a few static features are extracted from the vowel stable time-segments. Features include the fundamental frequency (F0), the first two formants (F1 and F2) and the vowel location in the vowel triangle [7]. Cepstral and LPC analyses are employed to extract the features using 20 msec Hamming windows with 50% time overlap. To compare the results to those of normal speakers, the nominal values of the features are measured using a control group of four normal speaking female children at ages 7-13, similar to patients.

In addition, we introduce two new features called Relative Energy and Relative Duration as measures of speaker's control on vowel production, voicing, and duration. It is shown that the two measures are consistently meaningful for both normal and impaired speakers.

3. STUDIES

3.1. Long-Term Study

In this study, recordings are done prior to, and three, six, and nine months after the operation, with the CI prosthesis always kept on during recordings. Figure 1 shows the fundamental frequency values for different patients, averaged over the three vowels. The columns marked with "PRE" and "POST" show the average F0 values pre-operation and nine months post-operation, respectively. As the figure shows, the fundamental frequency has on average decreased about 9% post-operation, which is consistent with other existing research results [2,5].

Another consistent feature of vowels is their location on the vowel triangle, drawn on an F1/F2 plane. Most CI researchers have employed the deviations of F1 and F2 formants from their normal values as a quantitative measure of speech quality. However, we suggest to measure the two dimensional distance of vowel formants (F1 and F2) from their normal values located on the vowel triangle. Figure 2 shows the normal formant values (of the control group) as well as the pre and post-operation values, averaged for all the patients. As this figure demonstrates, vowels have moved much closer to their normal values nine months post-operation . To quantify the time-course of the improvement, Figure 3 shows the two-dimensional distance of vowels from their respective normal locations on the vowel triangle, averaged for all patients and vowels. As it shows, the rms distance has monotonically decreased from a value of 629 Hz (per-operation) to 255 Hz nine months post-operation.

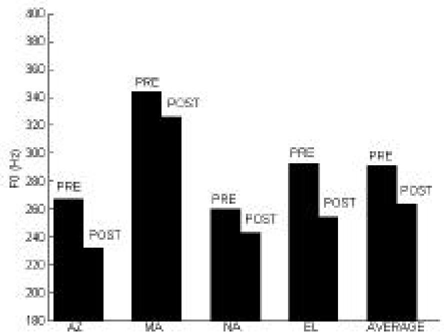

Figure 1: F0 (averaged over the three vowels) for four patients before (PRE) and nine months after (POST) operation, and also the averaged F0 over all patients and vowels.

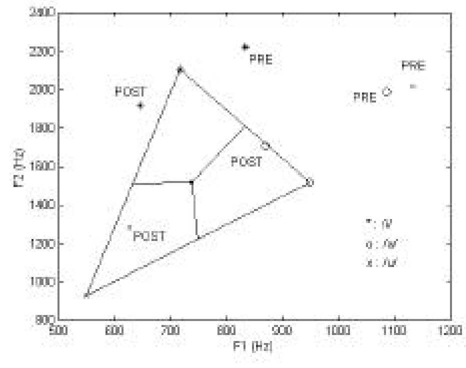

Figure 2: Averaged (over four patients) formant displacements for three vowels before (PRE) and nine months after (POST) operation, and for the normal speaker group (vowel triangle).

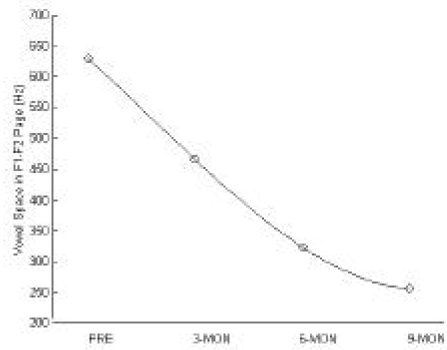

Figure 3: Averaged (over patients and vowels) rms F1/F2 formant distances (from respective normal formants) before (PRE) and after operation.

Another feature of the CI patients is the rather aspirated quality of their voice and the apparent weakness in their voicing. We suggest the use of the Relative Energy (RE), defined as the ratio of the vowel energy to the energy of its preceding consonant, and Relative Duration (RD), defined as the ratio of the vowel duration to the word (containing the vowel) duration. The motivation behind the RD features is the observation that often pre-lingual CI patients tend to abnormally lengthen the vowel parts of the word/syllable until their speech articulation organs get ready for uttering the following consonant(s). On the other hand, RE is obviously a measure of voicing strength. As the voicing becomes more regular and continuous, the average energy of the vowel (and hence RE) increases. For the RD feature, the ratio of duration of the second vowel to word duration in a set of CVCVC words were measured. For the RE feature, a set of CVC syllables were used. To examine the consistency of RE and RD, the ratio of standard deviation to mean values of the two features were measured for the control group. The two ratios (averaged for different vowels and speakers) were about 10% which clearly demonstrate the consistency of the features for normal speakers. Figure 4 and Figure 5 show the patterns of RE and RD variations by time (averaged for all patients and vowels), respectively.

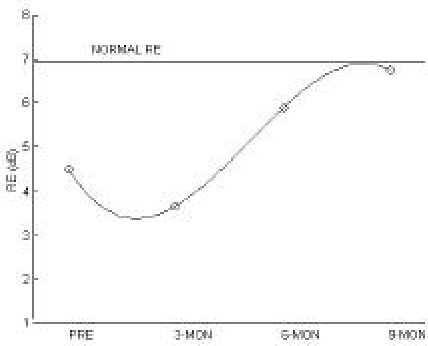

Figure 4: RE variations with time, averaged over patients and vowels, before (PRE) and after operation.

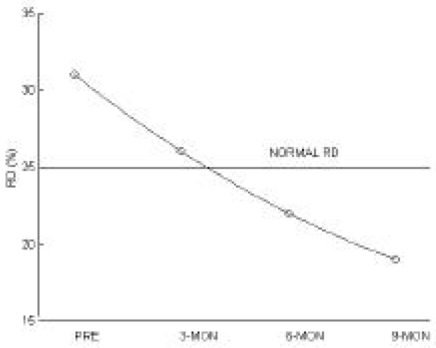

Figure 5: RD variations with time, averaged over patients and vowels, before (PRE) and after operation.

While the RE converges to a normal value by time, the RD feature converges to a value below normal. One possible cause of this observation could be the lack of proper stress pattern in patient's speech.

3.2. Prosthesis On-Off Study

The goal of this study is to experimentally evaluate the importance of the auditory feedback for the correct generation of vowels. At this stage, only static features of the three vowels /a/, /i/, and /u/ are considered. There have been few studies on the role of AF, reporting sometimes conflicting results (compare for example [1,2,5] to [3,4]). The approach chosen here is to repeat the same long-term study reported in the previous section (in PN condition), with an additional recording (at each session) after the patient has stayed in PF condition for at least 30 minutes, to remove any AF and short-term memory. Starting three months post-operation, a comparison is made between the speech features at PN and PF conditions, and then the study is repeated every three months to see the time-evolution of the results. Incorporating the time factor into the on-off studies is a crucial part of this work. As it will be shown, this engineering and time-dependent approach, can clarify some of the previous ambiguous results.

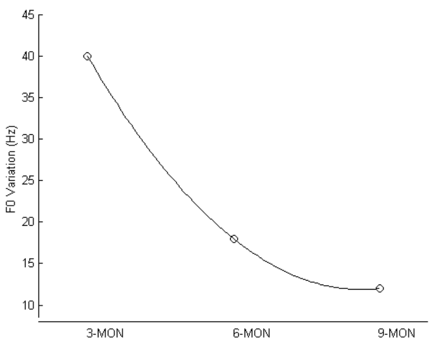

Figure 6: F0 variations (due to a PN to PF switch) with time after operation (averaged over patients and vowels).

Figure 6 shows the amount F0 would increase if the CI device is switched from PN to PF, at three to nine months after the operation. As shown, at three months post-operation, a PN-PF switch would cause an F0 increase of about 40 Hz on average. However, the same PN-PF switch would cause increases of only 18 and 11 Hz at six and nine months post-operation, respectively. Also Figure 7 demonstrates the rms displacement (in F1/F2 plane) of the first two formants when a PN-PF switch happens. Again, the tests are repeated every three months after operation. As shown, while at three month post-operation a PN-PF switch causes an rms formant displacement (degradation) of more than 250 Hz, the same PN-PF switch would displace the formants (in F1/F2 plane) less than 100 Hz. It can clearly be concluded from the two figures that *at least for the correct production of vowel static features, the dependency of the patients on the AF decreases with time after operation.*

To further validate the stated conclusion, the variations of the RE and RD features due to PN-PF switch are also evaluated. Figure 8 and Figure 9 show the results of such RE and RD measurements at three, six, and nine months after operation. Since at three months post-operation, most patients were not able to correctly pronounce the consonant preceding the vowel, the RE feature could not properly be evaluated, thus a NOT PRONOUNCED mark appears on Figure 8.

The results summarized in Figure 8 and Figure 9 also support the stated conclusion that patient's dependency on the AF, measured by vowel static features, decreases with time.

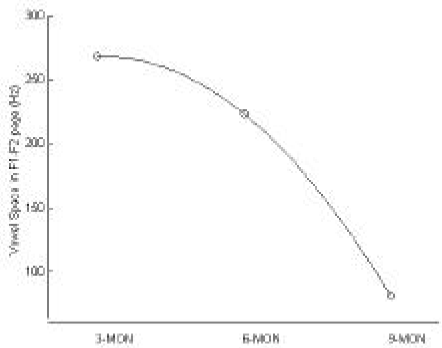

Figure 7: Rms F1/F2 displacements (due to a PN to PF switch) in time after operation (averaged over patients and vowels).

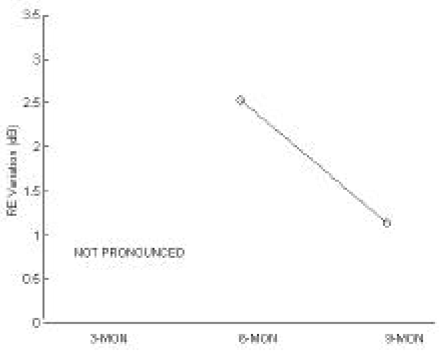

Figure 8: RE variations (due to a PN to PF switch) with time after operation (averaged over patients and vowels).

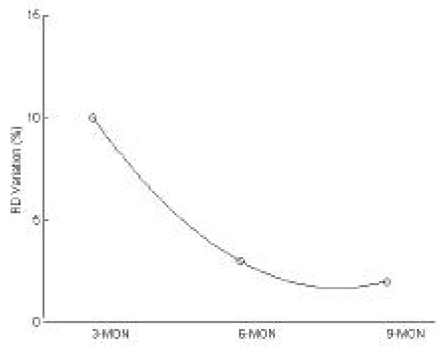

Figure 9: RD variations (due to a PN to PF switch) with time after operation (averaged over patients and vowels).

4. CONCLUSION AND FUTURE WORK

To be able to reach to conclusive and clear results, the scope of the studies were limited to include only vowel static features. Cochlear implanted children with pre-lingual deafness are among the best candidates for a study on the time-evolution of speech production skills. At the same time, by simply turning off the CI prosthesis, the effects of AF on human speech production can be investigated. To maintain objectivity, engineering and quantitative approaches were employed in this study to evaluate the speech quality. A control group of children, at ages similar to those of patients, helped to have a fair comparison to normal speaking people.

The new introduced RD feature has the advantage that, unlike the absolute duration of vowels used by previous researchers (see for example [2]), it is not much affected by rate of speech, and as the statistics presented in Section 3.1 show, is a consistent feature for a specific set of words.

Finally, instead of using single formant (F1 or F2) variations, as most researchers have done, we employed vowel location on the vowel triangle as a static feature. As it is well known [7], human listeners can discriminate most vowels based on the combined F1/F2 values. This justifies the use of vowel location on the F1/F2 plane as a static and consistent feature.

Quantitative results presented here show that:

1. Almost all of the static features considered have consistently improved in time, after the operation. It is particularly interesting to notice that due to the poor quality of the uttered speech, subjective evaluations could have hardly reached to conclusive results.

2. At least for the static features studied, the patient's reliance on the AF "decreased" consistently by time. While in the few first months after operation, loss of AF could degrade the speech quality to a noticeable extent, patient's dependency on AF decreased considerably after nine months.

By further continuing this research after nine months post-operation, one can obtain the ultimate convergence points of the static features and compare them to normal speaking people. We are now considering the static features of other classes of sounds such as short vowels, nasals and fricatives.

At the same time, we are extending the work to include more dynamic features of sounds such as formant transitions, voice onset time (VOT), and syllable and word prosodic features (pitch, loudness, and duration patterns).

It would be of great importance to compare the time-evolution of different static and dynamic features to each other, and to see how and why the AF affects each feature differently.

An intelligibility test (such as DRT) of the patient's speech by normal listeners would also give a quantitative evaluation of patient's overall speech quality. Also more long-term features of speech at phrase or sentence level (such as intonation pattern) have to be studied.

Acknowledgement

This work was done using facilities provided by the CI research center of AmirAlam hospital of Tehran. Authors wish to thank all the research team members, and specially professor M. H. Khalesi, head of the research center and the surgery team, and S. Abdi, the CI clinic director.

5. REFERENCES

1. **Tobey, E., Angelette, S., Murchison, C., Nicosia, J., Sprague, S., Staller, S., Brimacombe, J., and Beiter, A.,"**Speech production performance in children with multichannel cochlear implants," American Journal of Otology, 12(S)**,** 165-173, **1991**

2. **Svirsky, M., Lane, H., Perkell, J., and Wozniak, J.,**"Effects of short term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am., 92**,** 1284-1300, **1992**

3. **Tye-Murray, N., Spencer, L., and Gilbert-Bedia, E.,**"Relationships between speech production and speech perception skills in young cochlear implant users," J. Acoust. Soc. Am. 98, 2454-2460, **1995**

4. **Tye-Murray, N., Spencer, L., Gilbert-Bedia, E., and Woodworth, G.,**"Differences in children's sound

production when speaking with a cochlear implant turned on and turned off," Journal of Speech and Hearing Research, 39**,** 604-610, **1996**

5. **Matthies, M., Svirsky, M., Perkell, J., and Lane, H.,**"Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," Journal of Speech and Hearing Research, 39**,** 936-946, **1996**

6. **NUCLEUS USER MANUAL,** Cochlear Corporation, **1998**

7. **Deller, J., R, Proakis, J., G., and Hansen, J., H.,** Discrete-Time processing of speech signals, McMillan Publishing Company, **1993**