3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

# PERCEPTION AND PRODUCTION OF MOOD IN SPEECH BY COCHLEAR IMPLANT USERS

*David House*

Dept. of Linguistics and Phonetics, Lund University
Helgonabacken 12, S-223 62 Lund, Sweden

ABSTRACT

This paper presents results from a study where speaker mood identification performance of 17 cochlear implant users at two weeks following processor activation was compared to the performance of 13 and 11 members of the same group at 6 months and at more than one year after processor activation respectively. The cochlear implant users as a group showed difficulties confusing happy with angry and sad. Group performance improved only slightly over time while individual differences were considerable, ranging from 25% to 90% correct identification. Changes in production performance were also studied in three of the implant users. In the production study, the most dramatic improvements occurred in recordings made six months after activation. The length of time needed to regain voice control after deafness may indicate a difference between reacquisition speed in perception and production in cochlear implant users.

## I. INTRODUCTION

It has been shown that listeners with moderate to severe hearing loss have difficulties in identifying the mood of speakers in recorded test sentences [1, 2]. These listeners typically confuse happy (as the intended mood) with angry and sad. Cochlear implant users demonstrate a similar confusion pattern [3]. This paper addresses the question of whether implant users improve in their identification of speaker mood during the first year of implant use and what possible acoustic cues are used by these listeners.

Secondly, as an increasing number of postlingually deafened individuals receive cochlear implants, there has been a rising interest in the effect of implants on speech production [4, 5, 6, 7, 8, 9, 10]. This paper also addresses the question of whether implant users have similar difficulties in the production of different moods, and also if there is improvement during the first six months of implant use.

The use of different moods as speech perception and production material is also relevant in that it can serve as an evaluation tool for ascertaining perception of prosodic cues, and for studying changes in speech production particularly concerning control of fundamental frequency, intensity and voice quality. In the latter instance, production studies can be made comparing pre-implant speech (no self-hearing) with post-implant speech (after the reacquisition of self-hearing). This type of comparison can have implications for the understanding of the role of feedback in speech production [7, 8, 9, 11].

## II. PERCEPTION STUDY

### 2.1 Material

A standard mood test was designed for the purpose of testing the performance of cochlear implant users, for comparing individual performance over time, and for comparing performance of implant users to that of conventional hearing aid users. The standard test consists of four intended prototypical moods: angry, happy, sad and neutral and two different semantically neutral utterances with situations described which correspond to the four moods.

The two utterances are a sentence: 'Nu flyttar jag' (Now I'm going to move) and a number: 'Tvåtusenfemhundratio' (Two-thousand-five-hundred-ten ). For the sentence (Now I'm going to move) the following situations are described: *ANGRY* – telling friends about the landlord who has drastically raised the rent for an apartment in disrepair, *HAPPY* – telling friends about the acquisition of a wonderful new apartment, *SAD* – telling friends the news of a separation and its consequences, and *NEUTRAL* – reporting a change of address to an insurance company.

For the number (2510) the situations are as follows: *ANGRY* – the sum of an unexpected tax debt, *HAPPY* – the sum of an unexpected tax rebate, *SAD* – the loss of a purse containing the unexpected tax rebate, and *NEUTRAL* – the number of books in a library.

Recordings were made of a female speaker of Southern Swedish (a trained speech pathologist) who spoke the utterances in the prototypical moods. One token of each mood and each utterance was chosen for the test making a total of eight different stimuli. Each token was meant to clearly convey the intended mood using conventionalized expression without exaggeration. The test was intended to be relatively easy for listeners with normal hearing with the four moods being characterized by clear differences in intensity and fundamental frequency and by certain differences in spectral and voice source characteristics. Compared to the neutral versions of both utterances, the happy versions show a greater F0 range with higher maximum values and greater overall intensity. There is also evidence of a raising of F4 compared to all other moods, possibly as a result of a shortened oral tract due to lip spreading (smiling). The angry versions show a compressed F0 range with lower maximum values than both the neutral and happy versions and an even greater overall intensity than in the happy versions. Finally, the sad versions demonstrate particularly low overall intensity compared to all other moods with F0 contours similar to those of the neutral versions although the maximum values are lower. Also apparent is final creaky voice in the sad versions. More details of the stimuli are presented in [2].

### 2.2 Subjects

Eighteen university students and teachers with normal hearing served as a control group. 16 postlingually deafened listeners using the Nucleus multichannel cochlear implant and one listener using the Inneraid implant (Richards Medical Co.) participated in the test one to two weeks after processor activation (four to six weeks postoperatively). 13 members of the same group participated in the test 6 months after activation and 11 of these listeners were also tested at more than one year following activation.

### 2.3 Test configuration and procedure

A test tape was made in which each stimulus appeared six times. The test was divided into two parts, the first part consisting of the sentence and the second part consisting of the number. The test tokens were randomized within each part and presented with a five-second pause between stimuli. The tape was stopped between the two parts and after the first 12 stimuli in each part. The test consisted therefore of a total of 48 stimuli divided into four blocks. Four practice stimuli were presented before each part.

The test was presented via a Revox A77 tape recorder and a high quality loudspeaker. The test was administered individually to each cochlear implant user and to the controls as a group on one occasion.

Each subject was given a printed instruction sheet and an answer sheet with four mood options: angry, happy, sad and neutral. The test took about 15 minutes including instructions and pauses between blocks.

### 2.4 Results

Results for the control group showed 98% correct identification for the sentence (part 1) and 100% correct identification for the number (part 2). The cochlear implant users as a group showed the expected difficulties confusing angry as the intended mood most frequently with happy and neutral, happy as the intended mood primarily with angry, and sad as the intended mood primarily with neutral. Group performance is presented as confusion matrices in Tables 1 - 3. Group performance improved successively over time from 38% to 44% for the sentence and from 50% to 59% correct identification for the number. However, a chi-square test of independence showed the distribution differences over time for each part to be significant only for part two (the number) for *HAPPY* identification:

$$\chi^2 \text{ (4 df)} = 10.943 \ p<.05$$

and for *SAD* identification:

$$\chi^2 \text{ (4 df)} = 13.003 \ p<.05.$$

All other differences were found to be non-significant (p>.05). Individual differences were considerable, ranging from 25% to 90% correct identification.

### 2.5 Discussion

The patterns of confusion found in these experiments corroborate the main results of previous studies for both hearing-impaired subjects using conventional hearing aids and users of cochlear implants [1, 2, 3]. The confusion patterns form two main groups: happiness confused with anger and sadness confused with neutrality. This grouping would seem to reflect the use of intensity as a primary cue since higher intensity is common to both happiness and anger, while lower intensity is common to neutrality and sadness. It is clear, however, that many listeners also attend to F0 cues which is evident from responses of sadness and neutrality for anger as the intended mood. These three mood types have more similar F0 contours relative to the wide range of the F0 excursions for the happy mood. Here, an inability to perceptually integrate intensity and F0 may contribute to these confusions. Further difficulties may arise from an inability to perceive spectral differences in the frequency regions of F3 and F4 as these difference may supply important minor cues for identifying happiness. This could help explain the

**Table 1.** *Confusion matrices showing results of the two test utterances (sentence and number) for 17 listeners at 2 weeks after implant activation.*

| **Sentence (Part 1)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **46** | 16 | 14 | 24 |
| **Happy** | 45 | **27** | 19 | 9 |
| **Sad** | 6 | 10 | **38** | 46 |
| **Neutral** | 21 | 30 | 7 | **42** |
| **Total correct = 38%** | | | | |

| **Number (Part 2)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **50** | 18 | 5 | 27 |
| **Happy** | 52 | **32** | 12 | 4 |
| **Sad** | 5 | 6 | **62** | 27 |
| **Neutral** | 5 | 33 | 8 | **54** |
| **Total correct = 50%** | | | | |

**Table 2.** *Confusion matrices showing results of the two test utterances (sentence and number) for 13 listeners at 6 months after implant activation.*

| **Sentence (Part 1)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **47** | 17 | 13 | 23 |
| **Happy** | 50 | **26** | 20 | 4 |
| **Sad** | 0 | 9 | **50** | 41 |
| **Neutral** | 17 | 33 | 5 | **45** |
| **Total correct = 42%** | | | | |

| **Number (Part 2)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **41** | 28 | 7 | 24 |
| **Happy** | 51 | **45** | 0 | 4 |
| **Sad** | 1 | 3 | **83** | 13 |
| **Neutral** | 4 | 24 | 7 | **65** |
| **Total correct = 58%** | | | | |

**Table 3.** *Confusion matrices showing results of the two test utterances (sentence and number) for 11 listeners at 1 year after implant activation.*

| **Sentence (Part 1)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **48** | 17 | 15 | 20 |
| **Happy** | 39 | **39** | 13 | 9 |
| **Sad** | 2 | 3 | **47** | 48 |
| **Neutral** | 24 | 18 | 17 | **41** |
| **Total correct = 44%** | | | | |

| **Number (Part 2)** | *RESPONSE in %* | | | |
|---|---|---|---|---|
| *INTENDED* | **A** | **H** | **S** | **N** |
| **Angry** | **54** | 17 | 45 | 24 |
| **Happy** | 53 | **42** | 0 | 5 |
| **Sad** | 3 | 0 | **79** | 18 |
| **Neutral** | 2 | 29 | 8 | **62** |
| **Total correct = 59%** | | | | |

consistent low identification score for the happy mood. Implant listeners may, however, be able to perceive some voice source cues such as creaky voice which could aid in identifying the sad mood (see Tables 1-3).

For the group as a whole, time after activation did not significantly influence the results. Confusions present at two weeks after activation are maintained six months and one year following activation. Generally, then, it seems that implant users' perception of the cues needed for identifying speaker mood is relatively stable as early as two weeks after processor activation and corresponds to that of users of conventional hearing aids. A few individual subjects did, however, improve substantially after one year of implant use.

## III. PRODUCTION STUDY

### 3.1 Material and subjects

The same speech material and situations used for the perception study were also used for the production study (see section 2.1). The subjects were three of the first patients to receive a Nucleus multichannel cochlear prosthesis within the cochlear implant program at the Department of Audiology, University Hospital, Lund. All subjects are speakers of dialects of southern Swedish, two are females in their early 40's (S1 and S2) and one is a male in his late teens (S3). Prior to receiving their implants, S1 and S2 had been profoundly deaf for more than five years while S3 had been profoundly deaf for between 1 and 2 years. In conjunction with processor activation and adjustment, each subject received intensive voice training once a day for two weeks.

### 3.2 Recording procedure

The subjects were recorded in a sound-treated booth using a high quality professional tape recorder at the Department of Logopedics and Phoniatrics, University Hospital, Lund. Recordings were made shortly before surgery and at two weeks and at six months after implant activation. The subjects were given printed sheets explaining the situation for each mood and asked to read each utterance three consecutive times in each of the four moods.

### 3.3 Acoustic measurements

A fundamental frequency contour was extracted for each of the three repetitions of the four moods for each speaker using the LuPP program (Lund University) on a Macintosh II computer. Two intonation parameters were of particular interest in the analysis, namely F0 mean and F0 range, since these two parameters can be exploited in different ways to signal different moods [2]. While methods of quantifying these parameters are a matter of discussion [12], a method of approximating the parameter values using measured F0 maximum and minimum values in each utterance was employed. In several instances these values were obtained by measuring individual period lengths especially where voice source perturbations created problems for the F0 extraction algorithm.

For F0 mean, which corresponds to an approximation of the average fundamental frequency for the utterance in Hz, the arithmetic mean of F0 min and F0 max over the three repetitions was calculated for each mood. For F0 range, F0 max was divided by F0 min giving a 'range quotient'. This method has been used previously by Touati [13] for comparing individual accent patterns. This method clearly has limitations and would not be suitable for utterances containing several accents. However, for the short test sentence 'Nu flyttar jag' with only one accent, the method provides a quick and suitable approximation of both F0 mean and F0 range.

### 3.4 Listening tests

As a complement to the F0 analysis procedures presented above, formal listening tests were carried out using the second of the three repetitions of each mood produced by S1 prior to implant activation (Test A) and six months following implant activation (Test B).

To correct for possible order and learning effects, two independent groups of listeners served as subjects (Group 1, n=12 and Group 2, n=15). Both groups of listeners were comprised of beginning students in linguistics who voluntarily participated in the tests. The stimuli for Test A were divided into two blocks of 12 stimuli each. The stimuli were randomized within each block. Each token was repeated three times in each block. The same procedure was applied to Test B. The tests were presented to Group 1 in the following order: A B B A and to Group 2 in the reverse order, i.e. B A A B.

In a forced choice task, the listeners were asked to categorize each stimulus as one of four possible moods (angry, happy, sad or neutral). The tests were presented via a high quality tape recorder and loudspeakers. The listeners marked their responses on a prepared answer sheet.

### 3.5 Results

Table 4 shows F0 mean and F0 range for the three speakers and the four intended moods for three recording sessions. After two weeks, S1 lowers her F0 mean primarily for the neutral mood. After six months, however, she achieves a high degree of mood separation using F0 mean. Particularly noteworthy is the separation of happy and angry involving a substantial lowering of F0 mean for angry. For this speaker, no major change is seen in F0 range after two weeks. However, after six months, the speaker has clearly increased her F0 range for the happy mood.

Results for S2 are remarkably similar to those of S1. The greatest separation of all moods for both F0 parameters occurs six months after processor activation. As is the case for S1, F0 mean for angry and happy coincide after two weeks but are separated after six months as the speaker produces a higher F0 for happy and a lower F0 for angry. F0 range for happy also increases dramatically after six months.

**Table 4.** *F0 mean in Hz and F0 range quotient (F0 max/F0 min) shown in parentheses for four different moods produced by the three subjects on three different occasions.*

| SUBJ 1 | PRE-OP | 2 WEEKS | 6 MONTHS |
|---|---|---|---|
| **Angry** | 228 Hz (1.52) | 208 Hz (1.61) | 175 Hz (1.50) |
| **Happy** | 214 Hz (1.39) | 206 Hz (1.64) | 210 Hz (2.07) |
| **Sad** | 207 Hz (1.26) | 192 Hz (1.42) | 197 Hz (1.66) |
| **Neutral** | 202 Hz (1.40) | 164 Hz (1.41) | 164 Hz (1.52) |

| SUBJ 2 | PRE-OP | 2 WEEKS | 6 MONTHS |
|---|---|---|---|
| **Angry** | 300 Hz (1.53) | 261 Hz (2.02) | 229 Hz (1.79) |
| **Happy** | 305 Hz (1.57) | 258 Hz (1.63) | 304 Hz (2.43) |
| **Sad** | 206 Hz (1.13) | 226 Hz (1.44) | 214 Hz (1.58) |
| **Neutral** | 197 Hz (1.36) | 190 Hz (1.35) | 186 Hz (1.41) |

| SUBJ 3 | PRE-OP | 2 WEEKS | 6 MONTHS |
|---|---|---|---|
| **Angry** | 251 Hz (1.84) | 280 Hz (1.22) | 174 Hz (1.86) |
| **Happy** | 272 Hz (1.72) | 248 Hz (1.55) | 236 Hz (1.38) |
| **Sad** | 263 Hz (1.71) | 188 Hz (1.69) | 236 Hz (1.54) |
| **Neutral** | 206 Hz (1.55) | 193 Hz (1.35) | 141 Hz (1.52) |

Results for S3 are not as uniform as for the other speakers. He does, however, demonstrate a substantial lowering of F0 mean for neutral after six months and also produces a separation between angry and happy by lowering his F0 mean for angry.

Results from the listening test clearly show that S1 was dramatically more successful at conveying happiness in the post-implant recording. Sadness also showed an increase in identification. Anger, on the other hand, received a higher identification score in the pre-implant recording being confused with neutral in the post-implant recording. This corresponds very closely to the acoustic data presented in Table 4. A detailed discussion of the results of the listening test is presented in [10].

### 3.6 Discussion

The results of this study document a change in the use of fundamental frequency for three postlingually deafened speakers after the activation of their cochlear prostheses. The analyses of the production of different moods show the most dramatic changes occurring between two weeks and six months after activation. Here we can conjecture that voice training initializes a learning process which continues with the use of the implant. The expression of the moods involves a conventionalized use of intonation parameters and requires a considerable amount of control over fundamental frequency. It is also interesting to note that the two moods which coincide most often in the pre-operative and two-week production of S1 and S2 are happy and angry, two moods which hearing impaired listeners and implant users have considerable difficulty in differentiating perceptually.

The listening test, while only using production from one speaker (S1), clearly shows the increase in control of F0 from the pre-operative recording to the six-month recording. The dramatic improvement in expressing happiness is certainly noteworthy, and the social implications of the expression of this mood are not to be underestimated.

## IV. GENERAL DISCUSSION AND CONCLUSIONS

The results of the perception study suggest that implant users process the acoustic cues to mood by using intensity as the primary cue, fundamental frequency as a strong secondary cue and spectral and voice source characteristics as weak secondary cues. Perception and integration of some of the secondary cues can be seen as essential for discriminating between happy and angry. Some individual implant users are, however, able to integrate and interpret these cues quite successfully. Time after activation does not seem to be a major factor although some individual subjects did improve substantially after one year of implant use.

In the production study, some changes in the use of fundamental frequency to signal speech mood were apparent immediately after processor activation. However, the most dramatic changes in the use of fundamental frequency occurred in recordings made six months after activation where the speakers were able to produce differences between happiness and anger by varying mean F0 and F0 range. The length of time needed to regain voice control after deafness may indicate a difference between reacquisition speed in production and perception in cochlear implant users. Evidence suggests that subjects who have been deaf for many years lose laryngeal control successively over time developing deviant speech slowly resulting in habitual changes in motor control [11]. The fact that regaining a high degree of voice control after implant activation takes time is therefore not surprising. Perception, on the other hand, may rely on memory which, once activated, is relatively unchanged from the time before deafness. Thus listeners do not generally show a great amount of change in perception abilities between two weeks and one year, even though slight improvement can be noted.

In conclusion, the results of this study are encouraging for the users of cochlear implants. It is apparent that prosodic cues can be processed quite soon after activation and are used in much the same way as by hearing impaired listeners using conventional hearing aids. It is also clear that a combination of voice training and a cochlear prosthesis can result in considerable improvement in laryngeal control for postlingually deafened speakers.


## ACKNOWLEDGEMENTS

This research was carried out under a grant from the Bank of Sweden Tercentenary Foundation to the Department of Linguistics and Phonetics, Lund University. Recordings and speech training were carried out by Ursula Willstedt and Kajsa Johansson as part of the Cochlear Implant Program, Department of Audiology, Lund University Hospital. Sten Harris fitted the prostheses, and Bengt Almquist activated and adjusted the processors. Thanks are due to all the members of the Cochlear Implant Team for valuable discussions, the subjects for participating in the experiments, and the students for participating in the listening tests.


## REFERENCES


[1] A-M. Öster & A. Risberg, "The identification of the mood of a speaker by hearing impaired listeners," Speech Trans. Lab. QPSR 4, pp. 79-90, Stockholm, 1986.

[2] D. House, "On the perception of mood in speech: Implications for the hearing impaired," Working Papers 36, pp. 99-108, Dept of Linguistics, Lund University, 1990.

[3] D. House, "Cochlear implants and the perception of mood in speech," British Journal of Audiology, 26, p. 198, 1991.

[4] A-M. Öster, "Changes in speech with use of an implant," Speech Trans. Lab. QPSR 4, pp.13-22, Stockholm, 1988.

[5] G. Ball and A. Faulkner, "Speech production of postlingually deafened adults using electrical and acoustic speech pattern prosthesis," Speech Hear. Lang.: Work. Prog. (U.C.L.) 3, pp. 13-32, 1989.

[6] S. B. Leder and J. B. Spitzer, "Longitudinal effects of single-channel cochlear implantation on voice quality," Laryngoscope 100, pp. 395-398, 1990.

[7] J. Perkell, H. Lane, M. Svirsky, and J. Webster, "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. 91, pp. 2961-2978, 1992.

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

[9] A. Economou, and V. C. Tartter, "Speech changes following reimplantation from a single-channel to a multichannel cochlear implant," J. Acoust. Soc. Am. 92, pp. 1310-1323, 1992.

[10] D. House and U. Willstedt, "Changes in control of fundamental frequency and voice quality following cochlear implant activation and speech training," in Proc. Second International Conf. on Tactile Aids, Hearing Aids and Cochlear Implants, edited by A. Risberg, S. Felicetti, G. Plant and K. Spens, Dept. of Speech Comm. & Music Acoustics, KTH (Stockholm), pp. 201-210, 1992.

[11] H. Lane and J. W. Webster, "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. 89, pp. 859-866, 1991.

[12] E. Gårding, "Intonation parameters in production and perception". Proceedings of the Twelfth ICPhS, Vol. 1, pp. 300-304, Aix-en-Provence, France, 1991.

[13] P. Touati, Structures prosodiques du suédois et du français, Lund: Lund University Press, 1987.