Speech recognition of spectrums with 'holes' by children

Manasa Ranjan Panda & Asha Yathiraj*

Abstract

The identification of speech having holes in various bands in the spectrum was assessed in 30 normal hearing children. The speech material that was developed simulated perception through a twenty–two channel cochlear implant using noise band simulation technique. The 'holes' in the spectrum were created by filtering specific frequency bands which corresponded to the frequency bands of adjacent electrodes in a Nucleus cochlear implant. Responses were scored in terms of words and phonemes as well as consonants and vowels. It was found that there were less consonant errors than vowel errors for both word scoring and phoneme scoring. This error pattern was more with larger 'hole' size. It was also noticed that younger children in this study showed significantly poorer performance than older children.

Key words: spectral hole, filter, simulation.

Introduction

It is generally accepted that humans rely on cues that exist across several frequency bands to understand speech. The question of how listeners use and combine information across several frequency bands when understanding speech has puzzled researchers for many decades. It can be recognized even when the spectral information is reduced to three sinusoids that track the formant transitions over time (Remez, Rubin, Pisoni & Carrell, 1981). A high degree of speech discrimination and recognition is observed even under conditions of great reduction of spectral information. (Van Tasell, Greenfield, Logemann & Nelson, 1992; Shannon, Zeng, Wygonski, Kamath & Ekelid, 1995; Turner, Souza & Forget, 1995).

Understanding how speech is perceived after being processed through a cochlear implant is a challenge. In cochlear implants a relatively small number of electrodes activate tonotopic patches of neurons with a portion of the speech signal. Even with these abnormalities in physiologic process Shannon et al. (1995) found high levels of phoneme, word and sentence recognition could be achieved by adults with just four bands of information. This observation indicates how little is understood about recognition of speech under conditions of distorted spectral information.

Cochlear implantation is based on the idea that there are surviving neurons in the vicinity where electrodes are placed in the cochlea. The lack of hair cells and/or surviving neurons within the areas of cochlea essentially creates 'hole(s)' in the spectrum. The influence of the 'holes' in

* Professor of Audiology, All India Institute of Speech and Hearing, Mysore, India
e-mail: ashayathiraj@rediffmail.com

the spectrum in speech understanding is not well understood. It is not known whether the spectral 'holes' can account for some of the variability in the performance among cochlear implant users (Kasturi, Loizou, Dorman & Spahr, 2002). Hence, it is of interest to find whether recognition will be affected with the set of 'hole' pattern in speech.

Shannon, Galvin and Baskent (2001) assessed the impact of size and location of spectral 'holes' in cochlear implantees and normal hearing listeners. Results showed that holes in the low frequency region are more damaging than holes in the middle or high frequency region on speech recognition. Vidya, Rima and Yathiraj (2006) evaluated perception of seven lists of words having different band rejections and one list with no modification on thirty normal hearing adults. They found that despite information being filtered from the speech signal perception was not altered. Based on their findings they interpreted that as many as eight adjacent electrodes could be switched off in a cochlear implant without affecting speech identification. This gives an insight of how cochlear implant users perceive speech even when specific electrodes are switched off.

Various research findings contradict each other regarding number of electrodes recquired for better speech performance. Holmes, Kemker and Mervin (1987) evaluated speech perception of a patient fitted with a multichannel processor under different electrode conditions. Their study suggested that by increasing the number of programmed electrodes the subject's speech perception improved. This finding contradicts that obtained by others (Dorman, Loizou & Rainey 1997; Shannon et al. 1995; Turner, Souza & Forget, 1995). Dorman et al. (1997) processed vowels, consonants and sentences through software emulations of cochlear implant signal processors with 2-9 output channels. They found that high levels of speech understanding could be obtained using signal processors with a small number of channels. Despite many investigations the basic question of "how many electrodes for speech information" are still to be answered.

Knowing the relation between the numbers of electrodes activated and speech recognition is necessary for future designing devices as well as during counselling. Clinically a number of conditions may necessiate reducing the number of electrodes in a cochlear implant or not programming the electrodes. The effect of spectral resolution on speech recognition has received considerable attention in last few years. However, this attention is mainly concentrated on adults (Dorman et al., 1997; Fu, Zeng, Shannon & Soli., 1998; Holmes et al., 1987; Shannon et al., 1995, 2001; Vidya et al., 2006). It is theoretically and practically important to understand whether the limited spectral resolution is a key factor for children also. Eisenberg, Shannon, Martnez, Wygonski and Boothroyd (2000) reported that young children did not have sufficiently developed speech pattern recognition. Thus children may require more spectral channels than adults to obtain similar speech recognition skills. Hence there is a need to study to assess the effects of spectral 'hole(s)' in children using cochlear implants.

Besides evaluating the effect of spectral 'holes' on individuals using cochlear implant it has also been evaluated on normal hearing individuals using simulated material. This has been a

preferred method of study due to the ease with which the research can be conducted. It has also been found that controlling subject-related variables are a lot easier in a simulated condition. Thus the present study aims at investigating speech recognition in children using a cochlear implant simulated condition. Recognition of speech with varying spectral 'holes' will be studied.

Method

Participants

Thirty children with normal air conduction and bone conduction thresholds in the frequency range of 250 Hz to 8 KHz and 250 Hz to 4 KHz respectively, participated in the study. The children were in the age range of 7-10 years. They were native speakers of Kannada and were able to read and write the language. They had no history of any neurological disorders and had normal middle ear functioning as measured by tympanometry and acoustic reflexes.

Instrumentation

A Pentium IV computer with Matlab software was used for the development of the material. A CD burner (Nero 7 Ultra Edition) was used to write the material on a CD. To evaluate as well as present the test items a calibrated two channel audiometer (Madsen OB 922) with TDH 39 earphone was used. The middle ear status was determined with measurements from GSI Tympstar immittance meter. The speech material was played through a Philips CD player.

Material development

The speech identification test material "The Kannada Phonemically Balanced words" developed by Yathiraj and Vijayalakshmi (2005) was used to simulate speech processed through a cochlear implant. The test consisted of four lists of bisyllabic words with each list having 25 words. The CD recorded version of the test was used. The below mentioned procedure was used to simulate speech processed through a twenty-two channel cochlear implant.

The words of each of the original lists were randomized to create eight lists. These speech signals were band pass filtered into twenty-two frequency bands using 6^{th} order Butterworth filters. Crossover and center frequencies were calculated using the following equation relating the position on the basilar membrane to its characteristic frequency and assuming a basilar membrane length of 35 mm (Greenwood, 1990, cited in Rosen, Faulkner & Wilkinson, 1999):

$$\text{Frequency} = 165.4\ (10^{\,0.06x} - 1)$$

$$X = 1/0.06 \log (\text{Frequency}/165+1)$$

The envelope detection occurred at the output of each filter by full wave rectification and second order Butterworth low pass filter at 400 Hz. Forward and backward filtering were used to cancel the phase delays. These envelopes were then multiplied by signal correlated noise. Before being summed the signal was passed through an output filter similar to the analysis filter (Figure 1). Two conditions were created - "no 'hole' condition and "dropped" condition. In the dropped condition seven lists were created in which the output noise bands were simply omitted (band

pass filtered with 6th order Butterworth filter) from the processed signal. The 'hole' size was calculated corresponding to each of the dropped condition. The frequency bands of the band stop filters and 'hole' sizes are shown in Table 1. These band rejections were created in the frequencies which corresponded to frequency bands of groups of adjacent electrodes in a Nucleus cochlear implant. The above procedure was carried out using Matlab software. In the no 'hole' condition no band pass filtering was carried out. Prior to each test stimuli a 1 KHz calibration tone was also recorded. The altered stimuli were recorded on a compact disc (CD).

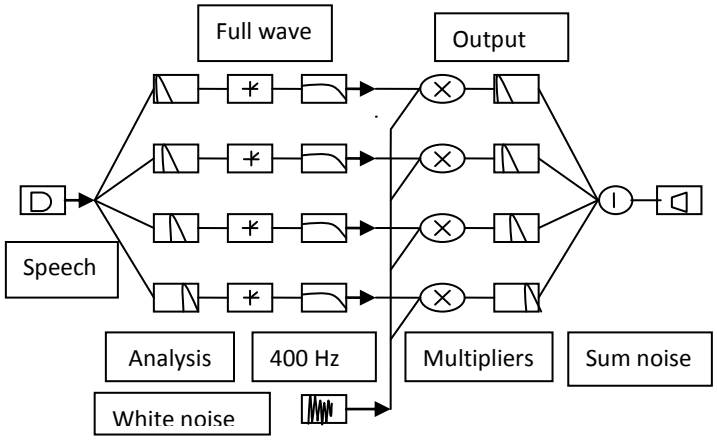

Figure 1: Block diagram of the processing used for transforming the speech signal

Test Environment

The test was carried out in a two-room audiometric set-up which was acoustically treated. The noise level was within permissible limits as recommended by ANSI (1991).

Table 1: Band rejection done for specific lists

Lists	**Frequency bands**	**Hole size**
LIST 1	438-813Hz	3.4 mm
LIST 2	938-1313Hz	2.0 mm
LIST 3	1563-2313 Hz	2.6 mm
LIST 4	2688-4063 Hz	2.8 mm
LIST 1 A	4688-6938 Hz	2.8 mm
LIST 2 A	438-1063 Hz	5.1 mm
LIST 3 A	438-1313 Hz	6.4 mm
LIST 4 A	No filter	-

Procedure

The hearing sensitivity of the participants was assessed using a calibrated two channel audiometer (Madsen OB 922) with TDH 39 earphone. A GSI Tympstar immittance meter was used to evaluate the status of the middle ear and to obtain acoustic reflex. Thirty children who met the subject selection criteria participated in the study.

The developed material was played using a Philips CD player. The output from the player was routed to the tape input of the two-channel audiometer. Prior to the speech signals being

presented the recorded 1 KHz calibration tone was played. This was used to adjust the VU meter deflection to zero. The signal was presented at 40 dB HL. The output from the audiometer was heard by the participants through circumaural headphone. Half the participants heard the signal in the right ear and the other half in the left ear. The order in which the lists were presented was randomized to avoid any list order bias. The participants were instructed to write down the words they heard.

Scoring

The written responses of the participants were scored in terms of phonemes as well as words that were correctly identified. For word scoring, each correct response was given a score of one and a wrong response a score of zero. The maximum possible word score for each list was 25. For phoneme scoring, each correct response was scored as one and wrong as zero. The maximum possible scores for list 1/1A, 2/2A, 3/3A and 4/4A was 103, 103, 104 and 106 respectively.

Results and Discussion

Statistical analysis was done using the SPSS software (version, 10.0). Descriptive statistics and repeated measure ANOVA was carried out on the data obtained from the thirty children to check significance of difference of speech identification scores across different filter conditions. This was done across all the filter conditions as well as non-filter condition. Both word scores and phoneme scores were analyzed for each filter condition.

A) Effect of different band stop filters on speech identification scores

The mean and standard deviation for the word scores and phoneme scores are shown in Tables 2. This information is provided for all eight lists that were evaluated. Details of the vowel and phoneme scores are discussed below.

Table 2: Mean and SD for the word and phoneme scores for different filter conditions

Lists	Band Rejections (Hz)	Word Score		Phoneme Score	
		Mean scores	SD of raw scores	Mean scores	SD of raw scores
List I	438-813	49.08% (12.27)	5.56	80.30% (82.73)	9.01
List II	938-1313	52.52% (13.13)	6.13	84.36% (86.90)	8.28
List III	1563-2313	56.28% (14.07)	6.02	83.46% (86.80)	10.00
List IV	2688-4063	54.00% (13.50)	7.16	81.00% (85.86)	11.51
List IA	4688-6938	57.48% (14.37)	5.85	85.95% (88.53)	8.53
List IIA	438-1063	58.28% (14.57)	5.67	87.01% (89.63)	7.62
List IIIA	438-1313	52.00% (13.00)	6.56	83.71% (87.06)	9.93
List IVA	No Filter	57.20% (14.30)	6.34	85.08% (90.13)	9.03

Note: Maximum word score = 25; Maximum phoneme scores vary from 103 to 106
Values given in bracket refer to the raw score

The word scores across filter conditions were relatively low, including the unfiltered condition. This indicates that the material simulating speech processed through a twenty-two channel cochlear implant resulted in reduced word scores without and with band stop filters. This highlights that the simulated material resulted in distortion reducing the overall intelligibility of the developed material.

A repeated measure ANOVA revealed that there was a significant difference between the word scores across the different lists [$F(7, 203) = 3.957$, $p < .05$]. The Bonferroni multiple comparison tests further indicated that for the word scores there was a significant difference between List I and List IIIA; List IV and List IVA at the level of 0.05 level of significance. Surprisingly, though the 'hole' was larger for List IIIA, the performance was poorer for List I. Though both the lists were equal in terms of phonemic balance and difficulty the variations in test items could have led to the difference in performance. The participants may have been able to utilize coarticulated information to a greater extent in List IIIA than in List I.

A significant difference was also found between List IV and List IVA. List IVA was without any 'hole' whereas List IV had a 'hole' of 2.8 mm. The cutoff frequencies used to create List IV were in the frequency region of the second format (F2) for several of speech sounds and F2 is a major cue to differentiate vowels (Carlson, Fant & Grantson, 1975). Hence removal of F2 information in List IV might have led to a significant difference compared to that of the no 'hole' condition. However, no significant differences were found for word scores across all other lists. Based on this finding it can be inferred that generally the listeners were able to combine the information across various frequency bands to perceive a "whole" signal.

The phoneme score across filter conditions was found to be significantly different on the ANOVA test across the eight lists [$F(7, 203) = 7.863$, $p < .05$]. This significant difference was observed between List I and Lists IA, III, IIIA, IV and IVA. In addition List IV significantly differed from List IVA. The difference in phoneme scores can be attributed to increase in 'hole' size. The 'hole' size for List I was larger (3.44 mm) when compared to List IA (2.8 mm), List III (2.6 mm) and List IV (2.8 mm).

The finding of the present study is in agreement with the findings of Shannon et al. (2001). They too reported that speech recognition decreased as the 'hole' size increased in normal hearing adults. They reported that a 4.5 mm 'hole' caused the performance to decrease significantly for consonants and vowel recognition. However, in the present study, it was found that a 'hole' size of more than 2.6 mm was able to reduce speech recognition. This difference of findings in both the studies might be due to test procedure and age of the participants. Shannon et al. (2001) carried out the experiment on adults in a free field condition while the present study was carried out on children under head phones. The task in the study by Shannon et al. (2001) was to identify medial vowels and medial consonants. However, in the present study, consonants and vowels in the initial, medial and final position had to be identified. This might have also attributed to difference in findings.

Shannon et al. (2001) also found that decrease in speech recognition was larger for apical 'holes' than basal 'holes'. In the present study also the 'holes' representing the apical region of the cochlea resulted in poorer scores (List I) when compared to the 'holes' representing the more basal regions. However, this was not seen for all 'holes' representing the basal region. No significant difference was found among other lists. It showed that listeners were able to effectively combine information from different frequency regions and perceive the speech signal despite the removal of certain frequency components.

In general, higher phoneme scores were obtained for phonemes in comparison to word scores. Similar results were also found by Olsen, Van Tasell and Speak (1997) in a group of normal hearing adults. In their study, phoneme scoring yielded scores that were on the order of 20% higher than scores for whole words heard. Barick (2006) also found a significant difference in word and phoneme scores. He recommended that word scores be calculated rather than phoneme scores since this scoring procedure depicts the perceptual problem better. However, he suggested that if the client was to be referred for auditory listening training the phoneme scoring procedure should be used.

Phoneme scores as a function of age was evaluated across filter conditions. All the participants were classified into four age groups, 7-8, 8-9, 9-10 and 10-11 years. The Duncan post hoc test revealed that the youngest group (7-8 year old) performed significantly poorer than the older three groups.

The finding regarding the performance of different age groups is similar to that reported by Eisenberg et al. (2000). They too noted that the youngest group in their study (5-7 year old) performed significantly poorer than their older children aged 10-12 years on a speech perception task. They had used age appropriate test material and found it on a variety of age matched tasks. They reported in their study that the younger children were more variable in their performance than adults and older children suggesting probable cognitive or task related factors playing a role.

Thus the findings of the present study substantiate the presence of a developmental trend in the perception of spectral 'holes'. This indicates that unlike the older children younger children are unable to carry out an auditory closure activity and guess the material that has been presented to them.

B) Phoneme errors as a function of different band stop filters

In addition to obtaining word and phoneme scores confusion of vowels and consonants were observed in each lists. Further, an error analysis was carried out for both the vowels and consonants to assess the error pattern. The analysis was done for three lists (List I, List III and List IA). These three lists were analyzed as they represented low (438-813 Hz), mid (1563-2313 Hz) and high (4688-6938 Hz) frequency band-stop filters respectively. Overall less consonantal errors were noticed than vowel errors (Table 3). Probably due to the larger number of redundant segmental cues present in consonants the participants were able to guess them despite the

presence of 'holes' in the spectrum. Thus if one cue is missed due to filtering listeners can perceive the other cues and identify the consonants. Despite vowels being more robust the number of redundant segmental cues present in them is less.

Vowel Errors

Among the three lists the least errors were observed in List IA, followed by List III and List I. In List IA, the cutoff frequency was between 4688-6938 Hz whereas the cues for vowel perception lie between 270 Hz and 2160 Hz (Peterson & Barney, 1954). Thus, in List IA, all the information for the perception of vowels was preserved. Hence the spectral 'hole' in it did not adversely affect the perception of vowel. In contrast the errors were maximum in List I as it removed the low frequency component in which first formants for all the vowels and second formants of many vowels lie (Peterson & Barney, 1954).

Further, the error analysis of these three lists revealed that there was confusion between short and long duration vowels. This was observed across all the three lists. This highlights that the simulated speech material affected the temporal cues required for the perception of long versus short vowels, resulting in this confusion. The error analysis for vowels also highlighted that in List I which contained the low frequency spectral 'hole', /u/ was confused with /i/ and /o/ was confused with /a/. Elimination of essential formant information probably resulted in this confusion.

Table 3: Consonants and vowel errors across different filter condition

List	Consonant errors (in %)	Vowel errors (in %)
List I	40.0	60.0
List II	44.0	56.0
List III	41.1	58.8
List IV	42.1	58.8
List IA	40.6	60.0
List IIA	46.6	53.3
List IIIA	35.3	64.7
List IVA	46.6	53.3

Consonant Errors

The consonantal error analysis was carried out in the same three lists as that done for vowels. In all the three lists voicing, place of articulation and manner was noticed to be affected. However, majority of errors were observed for place of articulation followed by manner and voicing errors. Manner and voicing cues have been found to be primarily temporal cues (Van Tassel, Soli, Kirby & Widin, 1987) and they require minimal spectral cues to be accurately perceived (Shannon et al., 1995). In contrast place cues require spectral cues which are affected by the spectral 'holes'. Similar results were also obtained by Shannon et al. (2001). They too reported that information received on place of articulation decreased considerably as the 'hole' size was increased particularly when the 'hole' was located apically. In the present study

maximum errors were obtained in List I which had a 'hole' located more apically. The 'hole' size in the List I was 3.4 which was larger than the other two lists. This apical location and larger size might have lead to more errors in List I. Also the alveolar /d/ was confused with labial /b/ and velar /k/ in List I and in List IA respectively. Generally the major cues for the perception of place of articulation of stops are the bursts and second format transition (Cooper, Delattre, Liberman, Borst & Gerstman, 1952). The spectal 'holes' in the speech material probably eliminated some of the major cues, causing confusion in the place of articulation perception.

Conclusion

From the findings of the present study it can be concluded that with increase in 'hole' size there was a deterioration in speech recognition scores. The apical location of 'holes' or band-stop filters affected speech perception more than the basal 'holes' or band-stop filters. However, not all the basal 'holes' had the similar adverse affect. The phoneme scores were higher in comparison to word scores. The error analysis indicated that consonantal perception was better than vowels for all filter condition. This error pattern was more with larger 'hole' size. However, this trend was not followed in all of the lists suggesting that listeners were able to combine the information from other frequencies to perceive the whole signal. Among vowels maximum confusion was noticed among short versus long vowels. For consonants more place of articulation confusion than manner and voicing confusion was observed. It was also noticed that 7-8 year old children showed significantly poorer performance than older children.

The present study will add to the current knowledge-pool of understanding speech pattern recognition in young cochlear implantees and their perceptual differences in the speech recognition with adults. Clinically findings of this study may help in predicting speech perception as a function of the electrodes that are switched on. Information from the study would be useful in counselling parents of young cochlear implantees or cochlear implantees regarding the speech sounds that will be affected if specific electrodes would be switched off.

References

ANSI. Maximum permissible ambient noise levels for audiometric test rooms, ANSI S3.1. New York: American National Standards Institute (ANSI), 1991.

Barick, S. K. (2006). High frequency-English speech identification test. Unpublished Master's dissertation submitted as part fulfillment for the degree of Masters of Science, to the University of Mysore, Mysore.

Carlson, R., Fant G. & Grantson B. (1975). Auditory Analysis and Perception of Speech. London: Academic Press Inc.

Cooper, F. S., Delattre. P. C., Liberman, A. M., Borst, J. M. & Gerstman. L. J. (1952). Some experiments on the Perception of Synthetic Speech Sounds. *Journal of Acoustical Society of America*, 24, 6, 597-606.

Dorman, M. F., Loizou, P. C. & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of Acoustical Society of America*, 102, 2403–2411.

Eisenberg, L., Shannon, R. V., Martnez, A. S., Wygonski, J. & Boothroyd, A. (2000). Speech perception with reduced spectral cues as a function of age, *Journal of Acoustical Society of America*, 107, 5, 2704-2710.

Fu, Q. J., Zeng, F. G., Shannon, R. V. & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of Acoustical Society of America*, 104: 505-510.

Holmes, A. E., Kemker, F. J. & Mervin, G. E. (1987). The effects of varying the number of cochlear implant electrodes on speech perception. *American Journal of Otology*, 8: 240-246.

Kasturi, K., Loizou, P., Dorman, M. & Spahr, T. (2002). The intelligibility of speech with 'holes' in the spectrum. *Journal of Acoustical society of America*, 112 (3), 1102-1111.

Olsen, W. O., Van Tassel, D. J. & Speaks C. E. (1997). Phoneme and Word Recognition for Words in Isolation and in Sentences, *Ear and Hearing*, 18, 3, 175-188

Peterson, G. E. & Barney H. L. (1954). Control methods used in a study of the identification of vowels. *Journal of Acoustical society of America*, 24, 183-

Remez, R., Rubin, P., Pisoni, D. & Carrell, T. (1981). Speech perception without traditional cues, *Science*, 212, 947–950.

Shannon, R. V., Zeng, F.-G., Wygonski, J., Kamath, V. & Ekelid, M. (1995). Speech recognition with primarily temporal cues, *Science*, 270, 303–304.

Shannon, R. V., Galvin, J.J. & Baskent, D. (2001): Holes in hearing, *Journal of the Association for Research in Otolaryngology*, 185-199.

Turner, C. W., Souza, P. E. & Forget, L. N. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *Journal of Acoustical society of America*, 97, 2568–2576.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987) Speech waveform envelope cues for consonant recognition. *Journal of Acoustical Society of America*, 82, 1152–1161.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J. & Nelson, D. A. (1992). 'Temporal cues for consonant recognition: Training, talker generalization and use in evaluation of cochlear implants'' *Journal of Acoustical Society of America*, 92, 1247–1257.

Vidya, M, Rima, D. & Yathiraj, A. (2006). Speech Identification of a Spectrum with Holes: Presented at ISHACON, 2006 held at Ahmedabad.

Yathiraj, A. & Vijayalakshmi (2005) The Kannada Phonemically Balanced Word Test developed in the department of Audiology, All India Institute of Speech and Hearing.