UNCLASSIFIED

AD NUMBER
AD837107
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies only; Administrative/Operational Use; 1968. Other requests shall be referred to Office Chief of Research and Development, Army, Attn: CRDSTI, Washington, DC 20310.**
AUTHORITY
OCRD D/A ltr, 14 Jul 1972

THIS PAGE IS UNCLASSIFIED

CASTELNOVO

EFFECTS OF SPECTRUM SAMPLING ON SPEECH INTELLIGIBILITY

ANTHONY E. CASTELNOVO
U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY
WASHINGTON, D. C. 20315

INTRODUCTION

The human performance experimentation research area at BESRL is concerned with behavioral functions common to a variety of Army jobs. Typical of this is the research in the Combat Communications unit which is concerned with finding ways to improve the performance of the human operator in military communications systems. One of the more serious problems the operator faces is that of noise which obscures the message. The noise may be broad band noise or appear in specific bands depending on the source. If the noise appears in relatively narrow bands these might be eliminated, however, there are no firm data on which we can estimate the effect on the operator's performance of using such a procedure. This study was designed to gain some preliminary information about the effect on performance of excising portions of the speech spectrum. It is recognized that sophisticated techniques, such as digital transmission of voice, are being developed and employed to overcome the effects of noise and to transmit communications in a secure form. There are, however, instances where these techniques are not feasible and it may be useful to reduce the amount of spectrum dealt with by excising selected bands. ( )

The frequency domain of speech and its relation to intelligibility has been the subject of research by many investigators (1,2, 3,4,5,6). These investigators have measured the average speech spectra and studied the effects on speech intelligibility of eliminating continuous bands from the upper and lower areas of the speech spectrum. Another way of treating the frequency domain is to remove bands simultaneously from several locations in the speech spectrum. Relatively little has been done to study the latter possibility. Kryter (7) noted that if several narrow bands are severely rejected, the intelligibility of the remaining spectrum is greater than that predicted by Articulation Index computations. In a follow up, Kryter (8) observed that for constant speech intelligibility the total "effective" bandwidth required for the best multiple pass band system is less than that required for the contiguous pass band systems by a factor of 2. This phenomena may be explained as a function of redundancy. Removing some narrow bands reduces redundancy though not necessarily intelligibility. That redundancy is a characteristic of the speech spectrum and that some reduction may be made without a corresponding reduction in intelligibility has been noted before. M. R. Schroeder (9) briefly reviews the work of Homer Dudley and notes his contribution to the origin of Vocoders which take advantage of the redundancy of the speech spectrum.

Apart from Kryter's work which employed a limited amount of filtering there appears to be no other relevant work in the literature. As for research on the effect of noise on a spectrum composed of discrete bands, there appears to be none at all. The present study is intended to gain preliminary information about the effect on intelligibility due to the number and size of segments which are excised and the effect of noise on the intelligibility of a speech spectrum composed of discrete bands. Such information would be useful in assessing the feasibility of eliminating segments of the spectrum which may carry particularly high levels of noise and for employing the spectrum space in the interstices for other uses.

PROCEDURE

To accomplish the sampling of the spectrum, a set of 24 electrical band-pass filters were used which permitted passing very narrow bands and which had very steep roll-off. A roll-off of 27 dB, from -3 to -30 occurred in the space of 20 cycles at the lower frequencies and in about 35 cycles at the higher frequencies. The bandwidths[1] of the individual filters at -16 dB varied from 50 to 115 cycles. The 24 filters allowed a total bandwidth of 1300 cycles, from 373 Hz to 1684 Hz to be passed. Each of the 24 bandpass filters could be switched in and out of the circuit independently of the other filters. The filter set was inserted in the system following a mixer which mixed the speech reproduced by an Ampex[2] tape machine with noise from a Grason Stadler[2] noise generator set for "speech" shaping.

---

[1] The bandwidth was computed for the -16 dB point to approximate the effective bandwidth of the filter. The -3 dB or -6 dB point often used in specifying the filter bandwidth does not take into account the intelligibility contributed by the filter skirt beyond that point.

[2] Use or mention of trade names is solely in the interest of precision of reporting procedures and does not constitute indorsement by USA BESRL or the Department of Army.

The combined stimulus material was amplified by a McIntosh[2] amplifier which was used to drive PDR-10[2] headsets. The system noise was -35 dB relative to the rms value of the speech (integrated over .3 seconds).

The stimulus material was presented to 36 test subjects. These were Army enlisted males under 30 years of age, hearing category I (a hearing test at the time of the experiment showed that all had excellent hearing) with no language problem and no previous experience in the communication field. The subjects were located in an Industrial Acoustics Company[2] series 1200 chamber which maintained a very low level of ambient noise.

The subjects were given 30 hours of training over a period of a week (6 hours a day for 5 days) in listening to PB words spoken by the three speakers. The materials had been subjected to filtering similar to that used in the experiment.

Following the training the subjects started the experimental sessions. These consisted of three one-half hour sessions on each of three days. Each half-hour experimental session was followed by a one-hour rest period. Six experimental conditions were presented each experimental period. Thus over the nine periods a total of 54 experimental conditions were presented. These were composed of 18 filter conditions and 3 noise conditions used with each filter condition. Figure 1 presents the filter conditions. Noise Condition 1 was zero noise from the noise generator, Noise Condition 2 was 25 dB below the maximum rms speech amplitude (integrated over .3 seconds). Noise Condition 3 mixed in noise at 15 dB below the maximum rms speech amplitude (integrated over .3 seconds).

The speech material consisted of eighteen Harvard PB word lists (10) which had been recorded by three talkers, six lists by each talker. Each subject was exposed to all 54 experimental conditions.

The design was such that all subjects, talkers, and word lists were associated with each experimental condition. The subjects had been instructed to respond to each stimulus word regardless of how unsure they were; and except for a very few instances a response was made to each word.

RESULTS

The data were reduced to the mean values for each of the 54 experimental conditions. These data are shown in Table 1. An analysis of variance was made for the main effects and for interaction of days,

---

[2] See footnote on page 2.

Table 1

COMPARISON OF EXPERIMENTALLY OBTAINED INTELLIGIBILITIES WITH THOSE COMPUTED BY USE OF THE ARTICULATION INDEX AT THREE NOISE LEVELS AT GIVEN BANDWIDTHS

Filter Config.	Band-Width	Obtain.	Comp	Obtain.	Comp	Obtain.	Comp
1	1311	59	68	51	58	37	30
2	931	59	44	43	33	29	18
3	917	58	42	43	33	29	18
4	1028	57	50	49	38	35	21
5	1022	55	50	44	38	30	21
6	903	55	42	39	33	29	18
7	906	55	42	43	33	30	18
8	750	51	35	37	27	26	15
9	817	50	40	40	30	29	16
10	739	50	35	37	26	25	14
11	688	46	32	40	25	25	14
12	667	46	30	40	24	26	12
13	577	43	25	26	18	15	11
14	478	38	20	30	15	19	9
15	600	37	27	30	20	20	11
16	815	34	35	18	26	11	14
17	49[illegible]	29	18	20	15	14	9
18	518	27	22	22	17	14	10

blocks of experimental conditions, period, speaker, filter condition and noise conditions. A small significant improvement was found over the three days of testing even though the experimental sessions had been preceded by a week of training. This, however, did not affect the results as each of the 54 experimental conditions were replicated 12 times on each of the three days. As was anticipated, the filter and noise factors were statistically significant beyond the .01 level. Blocks of experimental conditions, periods, and speakers produced non-significant F-ratios and the interactions of days by filter conditions and days by noise conditions showed a probability of occurrence of between .10 and .05. The filter by noise interaction was not significant although, as we note below, there is a significant change in the relationship between bandwidth and intelligibility as a function of noise level.

The intelligibilities produced by the different filter configurations were also compared to bandwidth and to expected intelligibility for a contiguous spectrum computed from the AI (Articulation

Index). (See Table 1.) Each of the filter configurations used were fairly representative of the total 1300 cycle band. The filter configurations were designed to have the same average AI per cycle as the total spectrum. This was done to maintain a linear relationship between AI and bandwidth (See Figure 2), and thus avoid confounding the variation in intelligibility resulting from the amount of spectrum with variations in intelligibility which might be caused by using spectra concentrated in particular areas of the spectrum. This was done even though for the area of the spectrum used in this study this was not very critical. The loss in articulation for the upper part of the spectrum, as compared to the lower part, is not great. This is reflected by the values for configurations 16 and 18 as shown in Figure 2.

Figures 3, 4, and 5 show the intelligibility data for the 18 filter configurations plotted for each noise level. The ordinate shows percent intelligibility and the abscissa the total bandwidth of the filter configuration. The data for each noise level was fitted (11) by a parabola of the form $Y = A + BX + CX^2$. Sixteen of the 18 configurations were included in the array fitted. The data points for configurations 16 and 18 were not included because they were not distributed samplings of the available spectrum. Configuration 18 included the lower 518 cycles of the spectrum and configuration 16 the upper 815 cycles approximately.

The redundancy in the speech spectrum is seen as the curvature exhibited by the data in Figures 3, 4 and 5. Figure 3, which presents the data for the lowest noise level, reaches a maximum of 58% for the full spectrum of 1300 cycles but appears to have reached an asymptote at about 1100 cycles; even at 1000 cycles the fitted curve does not show much loss. Also, the empirically obtained values for configurations 2, 3 and 4 for noise condition 1 are not significantly lower in intelligibility than that for configuration 1 (t-tests gave probabilities between .70 and .80). Figures 4 and 5, which present the data for increased levels of noise, show progressively less curvature. The curvature component for noise condition 1 is significant at the .001 level. For noise condition 2, the degree of curvilinearity is less and reaches significance only at the .05 level and the curvature for noise condition 3 appears slight and does not reach significance at the .05 level. On the basis of these comparisons it appears that there is in fact an interaction between noise level and filter configuration which was not apparent from the general interaction test.

If we compare configurations 16 and 18, which are examples of massed bandwidths with configurations 9 and 14, which are comparable in AI and bandwidth but which are distributed over the spectrum, we see that the distributed configurations produce a significantly higher level of intelligibility. (See Table 2.) This relationship holds for all three noise levels. Distributed configuration 4 has a bandwidth of 1028 cycles, which is substantially less than that of massed configuration 1, but the two give almost the same level of intelligibility at each of the three noise levels.

Table 2

RESULTS OF t-TEST COMPARISONS OF SELECTED MEANS

Filter Configurations	Noise Level Condition 1	2	3
1 vs 4	.72	.59	.66
9 vs 16	5.65[a]	8.40[a]	6.66[a]
14 vs 17	4.09[a]	3.64[a]	2.47[a]
14 vs 18	4.28[a]	3.20[a]	2.24[b]

[a]Significant at the .01 level
[b]Significant at the .05 level

Using the Articulation Index computed for each filter configuration and referring to the typical relationship between Articulation Index and Intelligibility of PB words (Figure 7 in Reference 4) the expected intelligibility was computed for these filter configurations and plotted in Figure 3 (shown as the dashed line). These values approximate a straight line. Using these computed intelligibilities as a reference we see that the experimentally obtained intelligibilities for the configurations with undistributed bandwidths, points 16 and 18, approximate what would be expected for this amount of bandwidth. The configurations with distributed bandwidths produce comparatively higher intelligibilities. There seems to be no apparent reason for the discrepancy between obtained and computed intelligibility for configuration 1.

Configurations 16, 17, and 18 which are the poorest (least well distributed) samplings of the spectrum also yield the lowest intelligibilities. As shown in Figure 1 they leave the largest areas of the spectrum unsampled. Configurations 16 and 18, as has been noted, are each composed of a single pass band. Configuration 17 consists of two pass bands, one at each end of the spectrum, which leaves a gap of 820 cycles in the center of the spectrum.

CONCLUSIONS

Segments of the speech spectrum may be excised under conditions of low noise without incurring a proportionate reduction in intelligibility as we might have expected for this spectrum had we used high or low pass filtering. Differences in size and number of segments excised, if the size of the excisions is not large, do not appear to differentially affect intelligibility. We see in Figures 3, 4, and 5 that configurations with different sized segments excised but with approximately equivalent bandwidths are of approximately equal intelligibility. This is best illustrated by the cluster formed by configurations 2, 3, 6, and 7 which are composed of different numbers of samples and have had different sized segments excised. However, if the size of the excision becomes a relatively large contiguous portion of the total spectrum, as is the case with configuration 17, excising several smaller segments, as was done in configuration 14, results in a significantly higher level of intelligibility for a given total amount of bandwidth. Thus, it appears that bands of about 200 cycles may be excised with no greater loss than would result from excising an equivalent amount as a larger number of smaller segments. Under some conditions relatively large bands of the spectrum may be excised without a significant loss in intelligibility.

It seems reasonable to expect that some gain in intelligibility may be made by excising narrow bands of the spectrum which have a poor speech-to-noise ratio. The human auditory system can be characterized as a filter with relatively slow roll off, that is, a sound from one part of the spectrum can interfere with the perception of a sound from another part of the spectrum. This spread of masking, as it is called, can cause relatively narrow bands of noise from one part of the spectrum to interfere with the perception of speech which lies in other parts of the spectrum. Future studies will be concerned with the effect on the operator's performance caused by removing portions of the spectrum which carry high levels of noise and consequently give rise to masking of adjacent spectrum areas.

REFERENCES

1. Licklider, J. C. R. and Miller, G. A. The perception of speech in S. S. Stevens (ED.), The Handbook of Experimental Psychology. New York: John Wiley and Sons, Inc. 1951. Ch. 26, Pp. 1040-1074.

2. Fletcher, Harvey. Speech and hearing in communication. Princeton, N. J.: D. Van Nostrand Company, Inc. 1953. Ch. 18, Pp. 418-422.

3. French, N. R. and Steinberg, J. C. Factors governing the intelligibility of speech sounds, Journal of the Acoustical Society of America, 1947. 19, 90-119.

4. Kryter, K. D., Flanagan, Gail and Williams, Carl. A test of the 20-band and octave-band methods of computing the articulation index. Contract AF 19 (604)-4061. Bolt Beranek and Newman, Inc. Cambridge, Mass. 1961.

5. Pollack, Irwin. Effects of high pass and low pass filtering on the intelligibility of speech in noise, Journal of the Acoustical Society of America, 1948. 20, 259-266.

6. Beranek, Leo L. The design of speech communication systems. Proc. IRE 35, 1947. 880-890.

7. Kryter, Karl D. Journal of the Acoustical Society of America, 1957. 29, 1262.

8. Kryter, Karl D. Speech bandwidth compression through spectrum selection, Journal of the Acoustical Society of America, 1960. 32, 547-556.

9. Schroeder, M. R. Vocoders: Analysis and synthesis of speech. Proc. IEEE 54, 1966. 720-734.

10. Egan, James P. Articulation testing methods. Laryngoscope, 1948. 58, 955-991(a)

11. Snedecor, George W. Statistical methods. Ames, Iowa: The Iowa State College Press, 1940. Ch. 14, Pp. 313-317.

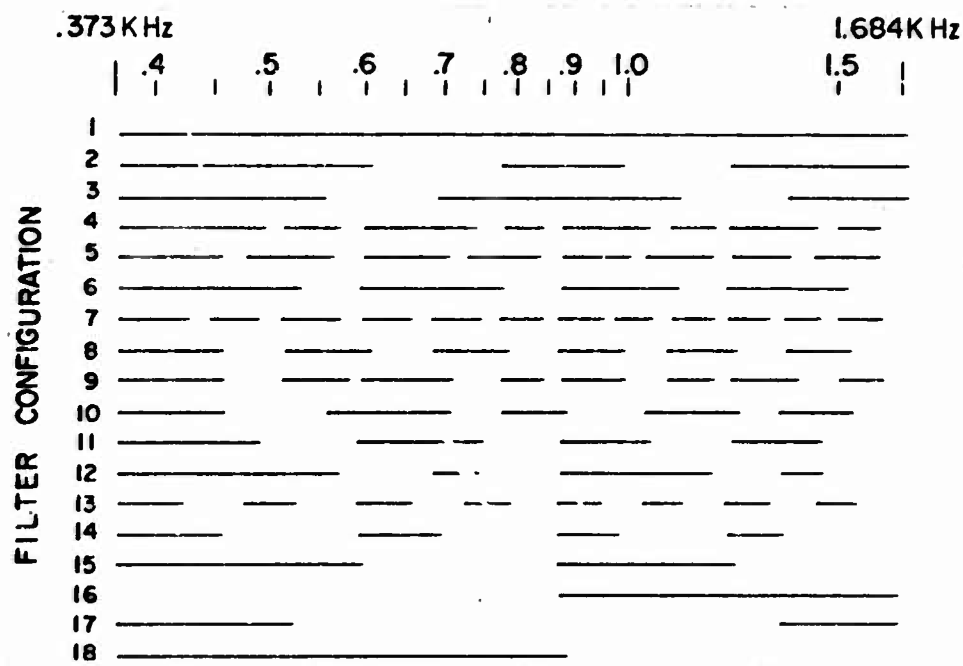

Figure 1. Representation of the filter configurations. (Lines indicate pass bands)

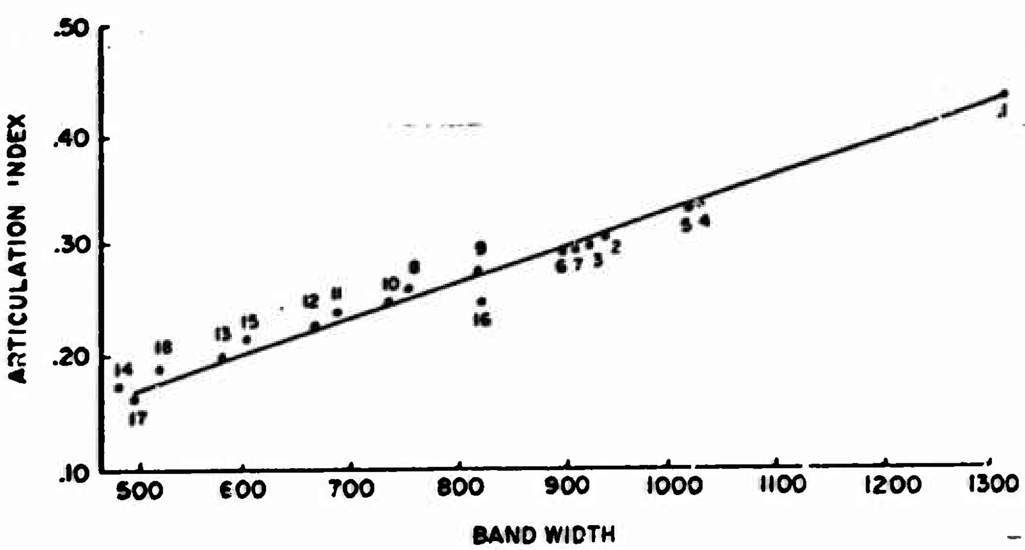

Figure 2. Relationship between Articulation Index and total bandwidth. (The data points are the filter configurations)

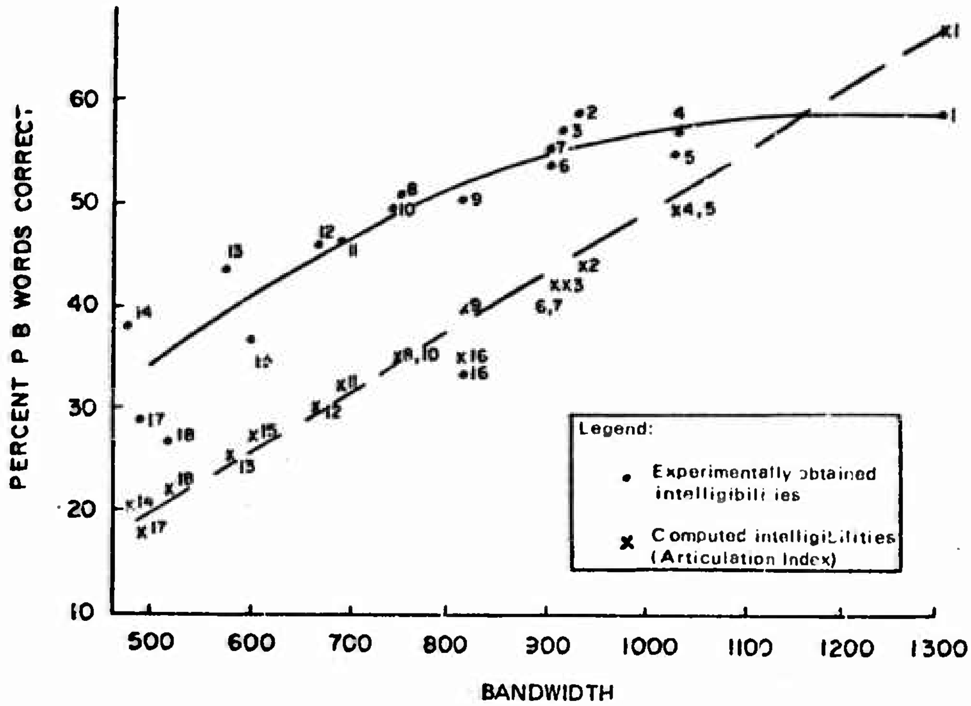

Figure 3. Comparison of experimentally obtained intelligibilities for noise condition 1 with computed intelligibilities

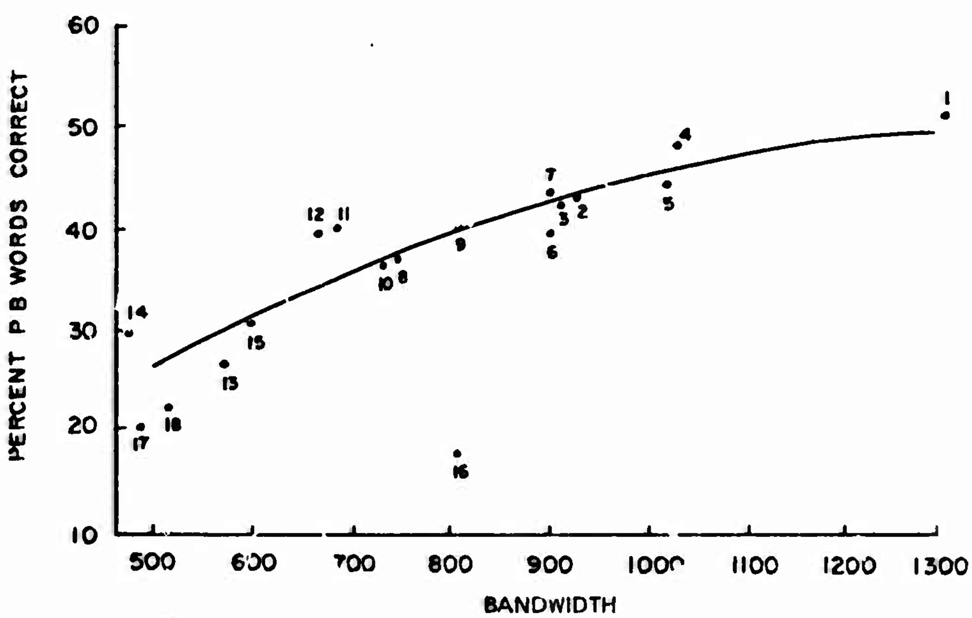

Figure 4. Experimentally obtained intelligibilities for noise condition 2

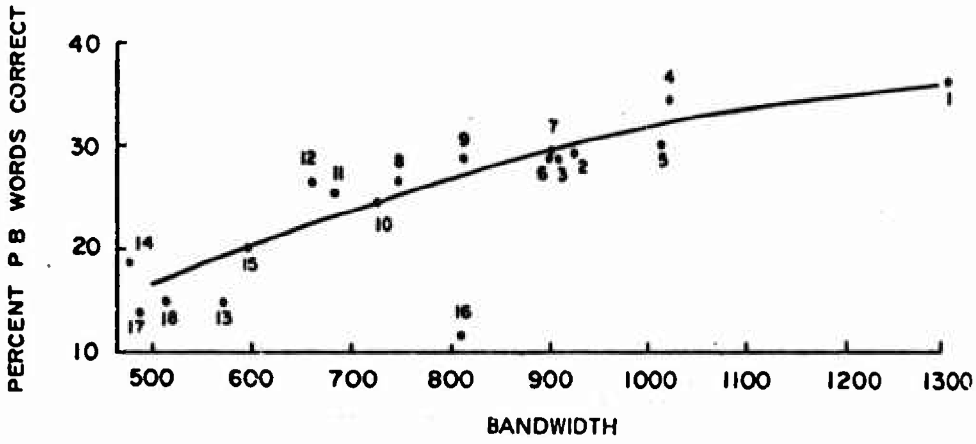

Figure 5. Experimentally obtained intelligibilities for noise condition 3

129