Experiments with Filtered Speech and Hearing-Impaired Listeners

David House

ABSTRACT

Speech perception experiments using LPC pitch-edited low-pass filtered speech stimuli were presented to normal and hearing-impaired listeners in an attempt to study the interplay of two kinds of frequency movement (Fo and formant transitions) in the perception of stops and CVC words in a speech context, and to test the hypothesis that increased movement of Fo can aid stop identification. The results indicated little difference in stop identification relating to Fo movement. There were, however, certain considerable differences in results between listeners with normal hearing and listeners with a noise-induced hearing loss indicating the possibility of an acquired perception stratgegy used by the hearing-impaired. Correlation between filter frequency, hearing loss and test performance was also observed.

1. INTRODUCTION

Intonation has been shown to be a major factor in determining the perception of syllable stress (Fry, 1958). Furthermore, it appears that fundamental frequency movement in sentence intonation provides an overriding cue to stress, and that this movement may produce a perceptually all-or-none effect, the important factor being movement per se and not magnitude of movement (Lehiste, 1970). Focus of a sentence, however, as defined by Jackendoff (1972) as being the "new" information not

shared by the speaker and the hearer, can be signaled by introducing emphatic sentence stress and thereby significantly increasing the movement and range of Fo (House, 1983).

Although intensity and duration are correlates of stress, fundamental frequency movement seems to provide relatively stronger perceptual cues to the location of stress (Lehiste, op.cit.). Fo movement then provides, from the point of view of dynamic perception (Johansson, 1975), a change in frequency over time which could be registered as an event by the perceptual mechanism. This event would in turn sharpen attention and aid in short-term memory retrieval of spectral cues. Since resolution of spectral cues can be seen as more crucial in the "bottom-up" processing of new information than in presupposive "top-down" speech processing (Marslen-Wilson & Tyler, 1980), it would be interesting to investigate to what extent segmental resolution could be facilitated by varying degrees of fundamental frequency movement where semantic focus is constant in "bottom-up" processing using semantically non-redundant speech stimuli.

Cutler (1976), using phoneme-monitoring reaction time experiments, concluded that a significant difference in reaction times can be attributed to the prediction of upcoming stress locations even when the target-bearing word in stress position does not contain a high-stress intonation contour. Reaction times were, however, substantially shorter when the target-bearing words contained high-stress contours.

Intonation, then, points toward the focussed word where segmental resolution is facilitated by virtue of sharpened attention on the part of the listener coupled with changes in pitch, vowel duration and intensity. The question raised here is one concerning the role of frequency movement in segmental resolution. Can the same movement in frequency used to signal location of sentence stress be used by the perceptual mechanism in interaction with other cues to aid in segmental resolution?

One possible interaction could be between pitch movement realized as frequency movement of harmonics of the fundamental in the vowel and vowel formant transition movement realized as resonance induced amplitude shifts between successive harmonics.

Since formant transitions are important cues for stop identification, such interaction might facilitate perception of transitions and aid in stop identification. On the other hand, interaction might not necessarily result in an amplification of formant transitions and would therefore not facilitate stop identification.

A further question that arises is that if these two kinds of movement in frequency interact to facilitate segmental resolution, what are the possible implications for listeners with hearing disabilities? It is well documented that individuals with moderate sloping sensorineural hearing losses have difficulty in identifying place of articulation especially in voiceless stops. However, subjects having similar audiometric configurations can differ radically in their performance in both synthetic and natural speech tests (Van de Grift Turek, et al. 1980; Risberg & Agelfors, 1978; Picket, Revoile, & Danaher, 1983). Could Fo movement as a correlate of stress aid hearing-impaired listeners in identifying stop consonants?

In attempting to answer these questions, word identification tasks could be presented through a filter roughly corresponding in frequency to a typical noise-induced audiometric configuration for hearing-impaired listeners. By presenting the stimuli both to listeners with normal hearing and to hearing-impaired listeners, the experiment would have two goals: 1.) to test frequency movement interaction as an aid to stop consonant identification in filtered speech, and 2.) to compare performance of listeners whose audiometric configurations correspond to the filter frequencies used in the

presentations. The latter goal might also help in exploring the relationships between perception of the speech wave filtered before reaching the auditory perifery and perception of the speech wave altered by an impairment of the auditory perifery.

2. METHOD

A. Linguistic material and stimuli

There is much debate concerning differences in perception of sense vs. non-sense speech utterances and utterances in and outside of a sentence frame in listening experiments (Pastore, 1981; Johnson & Strange, 1982). It seems, however, that when dealing with stimuli involving both sentence intonation and local segmental cues and their interaction, the closer the stimuli can be to real-life speech, provided of course that variables can be sufficiently controlled, the more we can learn about the communicative aspects of speech perception. (See also Gårding, 1967 for differences between juncture perception in sense and non-sense words.)

The carrier sentence "de' va' _ ja' sa'" (It was _ I said.) was selected such that a vowel would immediately precede the target word. 26 single-syllable CVC words were chosen having an initial voiced or voiceless stop (Table 1). Target word initial stop was considered as the perceptual target phoneme and Fo movement in the preceding and following vowel was to be altered to test interaction between stop identification and frequency movement in the two adjoining vowels.

Five of the test sentences were recorded by a male speaker of Southern Swedish in two versions, first with emphatic high stress given to the target word, then with indifferent low stress. The 26 test sentences were then recorded by the same speaker using neutral intonation and stress.

The five low-stress and five high-stress tokens were digitized

	TAL (speech)	KAL (bare)	BAL (dance)	DAL (valley)	GAL (to crow)
PAR (pair)	TAR (take)	KAR (tub)	BAR (bare)		
	TUR (luck/turn)	KUR (cure)	BUR (cage)	DUR (major key)	
	TÅR (tear/toes)	KÅR (corps)	BÅR (stretcher)		GÅR (go/walk)
PÅG (boy)	TÅG (train)		BÅG (cheat)		
	TAM (tame)			DAM (lady)	GAM (vulture)
	TOK (fool)	KOK (potful)	BOK (book)		

Table 1. Target words used in the stimulus sentences.

using a VAX computer at a sample rate of 10,000Hz. The tokens were then analyzed with a linear prediction analysis method (ILS program package). The five contours for the two intonation types were averaged to serve as a natural model for pitch editing of the sentences with neutral intonation. The 26 neutral intonation sentences were digitized and analyzed in the same manner.

Two stimulus versions of each sentence were then synthesized from the linear prediction coefficients with the pitch contour selected to conform to the low and high-stress intonation models respectively (Fig. 1). The resulting 52 stimuli were randomized in the computer. A pilot test using four listeners was run to determine a satisfactory low-pass filter cut-off frequency which would allow correct initial stop identification of around 50% in the target word. An eight order Butterworth low-pass filter with the cut-off frequency set at 900Hz, -48dB/octave allowed the 50% correct identification in the pilot study.

The stimuli were recorded on tape through the filter in thirteen blocks each containing four stimuli with a 10-second interval between stimuli and a 20-second interval between blocks. An additional block was placed at the beginning as a practice buffer.

B. Subjects

25 beginning speech pathology students at Lund University with normal hearing and unfamiliar with the test material participated in the experiment as part of a course requirement. 13 patients at the Department of Otorhinolaryngology, University Hospital, Lund, with noise-induced hearing losses of roughly similar audiometric configuration (beginning of slope at around 1000Hz, approximately -40dB at 2000Hz and -60dB at 3000Hz) voluntarily participated in the experiment as an extended part of routine audiological examinations. Patients were selected on the basis of previous pure-tone audiograms

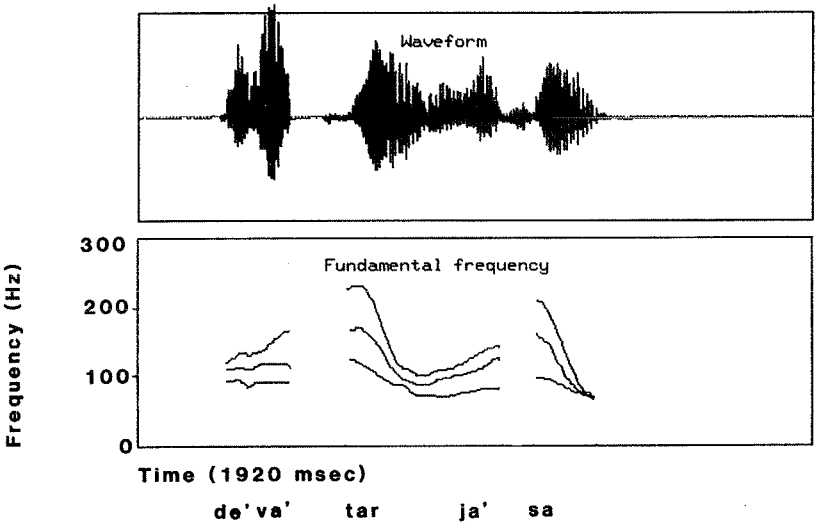

Figure 1. Waveform and fundamental frequency contours of one stimulus sentence. The middle contour represents the original neutral stress, the upper contour is the high-stress edited version and the lower contour is the low-stress edited version. The waveform represents the neutral version. Only the two edited versions were used in the test.

where relative symmetry of hearing loss was a criterion as well as break frequency and degree of slope.

C. Procedure

The normal-hearing subjects were tested in three groups on three different occasions. Written instructions and answer sheets were handed out and the instructions were read aloud by the experimenter. The subjects were informed that they would hear the carrier sentence "Det var _ jag sa." presented through a filter and that they should try to write the word following "var" in each sentence. If they were unsure they were requested to guess the word closest to the sound they heard. After the practice block was run subjects were allowed to ask questions. Testing took place binaurally in a sound-treated perception laboratory using a Revox A77 tape recorder and Burwen PMB6 Orthodynamic headphones. Sound level was checked as comfortable during the practice block. The test took 15 minutes.

The hearing-impaired persons were tested individually while sitting in a sound-insulated room. Routine pure-tone and speech audiograms were first made after which Békésy sweep-audiograms with pulsed stimuli were performed for each ear. This was done to more closely define the steepness and frequency location of the hearing loss. The same instructions were presented to the listeners except that they were asked to repeat the words orally instead of in writing. The responses were monitored outside the sound-insulated room over a loudspeaker and recorded on tape.

The stimuli were presented monaurally through a Revox A77 tape recorder, via the speech channel of a Madsen Clinical Audiometer Model OB70 and matched TDH-39 headphones with MX-41/AR cushions. The stimuli were presented at most comfortable level established during the practice block and generally corresponding to 30dB over speech threshold and to the level for maximum speech discrimination established during speech audiometry.

Monaural presentation was deemed advisable since hearing loss was not completely symmetrical. A second randomized version of the tape was made, and learning effects were minimized by presenting half the first version to the left ear, the entire second version to the right ear and then the remaining half of the first version to the left ear. After hearing the filtered stimuli the subjects were given a break and then the test was repeated using non-filtered stimuli. The test, including the Békésy audiograms but excluding the routine audiograms, took approximately an hour and a half. The patients were extremely cooperative especially considering the length of the test.

3. RESULTS

A. Normal-hearing listeners

All three groups showed similar patterns of initial stop identification for the filtered stimuli. Roughly one-half of the stops were correctly identified, and a general bias toward labials was observed. The voiced-voiceless distinction was perceived by all subjects in nearly all target words with place-of-articulation for voiced stops being somewhat easier to identify than for voiceless stops. The mean number of correct stop identifications for the normal-hearing subjects as a group (Group 1, Fig. 2) was higher (14.5 of 26) when the sentence and target word carried the indifferent, low-stress intonation contour than when the sentence and word carried emphatic, high stress (12.7 of 26). The difference was significant, $p<0.05$, running contrary to the hypothesis. Correct word identification, however, did not reveal a significant difference between low and high stress, although more low-stress words were correctly identified. Labials /p,b/ and the voiced velar /g/ were favored by the normal-hearing listeners (Fig. 3). In only one phoneme /d/ were substantial differences observed relating to stress contours. The low-stress versions received more than twice as many correct responses compared to their high-stress counterparts. The

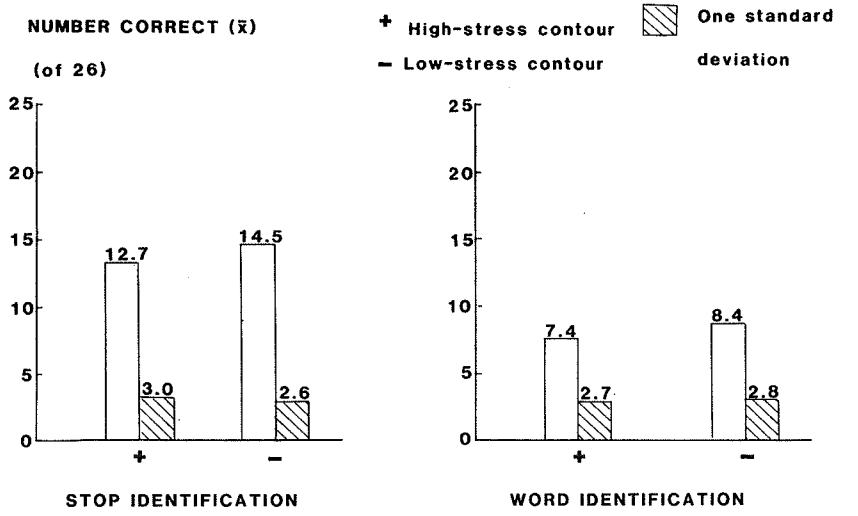

Figure 2. Mean number of correct identifications for stops and words in low-pass filtered sentences with high and low-stress fundamental frequency contours. (Group 1, listeners with normal hearing.)

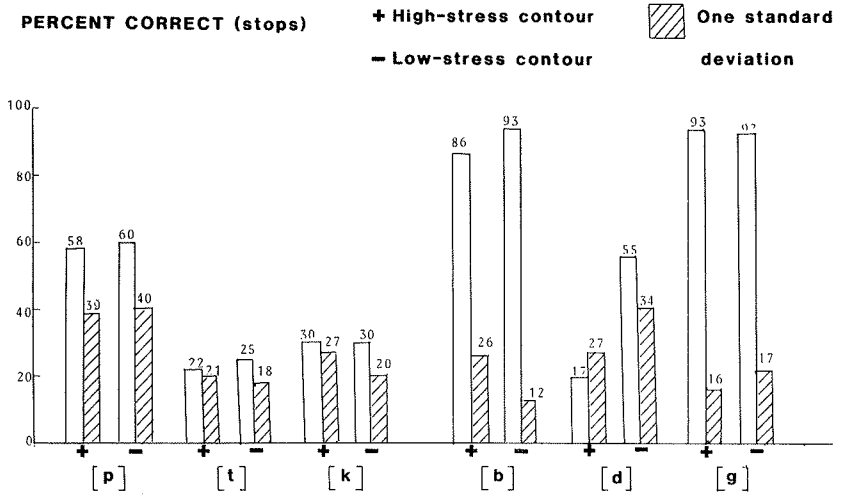

Figure 3. Correct initial phoneme identification in percent for Group 1 (Normal hearing)

vowels were nearly always correctly identified.

B. Hearing-impaired listeners

The number of stops and words correctly identified by the hearing-impaired listeners as a group was about the same as for the normal group: about one in two for stops and one in three for words (Fig. 4). Again, for both words and stops, identification was slightly better for sentences having the low-stress Fo contour, although here this difference was not significant, $p>0.05$. Labials and velars were again favored in the voiced stimuli results, but the dental phoneme /t/ was favored in the voiceless results (Fig. 5). Standard deviation for both stops and words in both stress categories was greater for the hearing-impaired group than for normals. As with the normal group, the vowels were nearly always correctly identified.

C. Differences between the two groups

As previously mentioned, the hearing-impaired group as a whole did better on both stop and word identification in the filtered speech, although the difference was not significant, $p>0,05$. There was, however, a striking difference between the two groups manifested by the preference for the voiceless dental /t/ among the hearing-impaired group which contrasted to the labial preference /p/ among the normal group.

Reactions to the test by subjects of the two groups also differed. Members of the normal-hearing group felt that the test was extremely difficult and frustrating. There were, however, substantial differences among members of the hearing-impaired group in both performance and reactions to the test. These listeners basically fell into two categories. Either they reacted much like the normal group complaining about the difficulty of the task and obtaining results similar to the normal group or they made many correct responses from the very beginning of the test, performed better throughout the test than the other groups, and did not feel that the

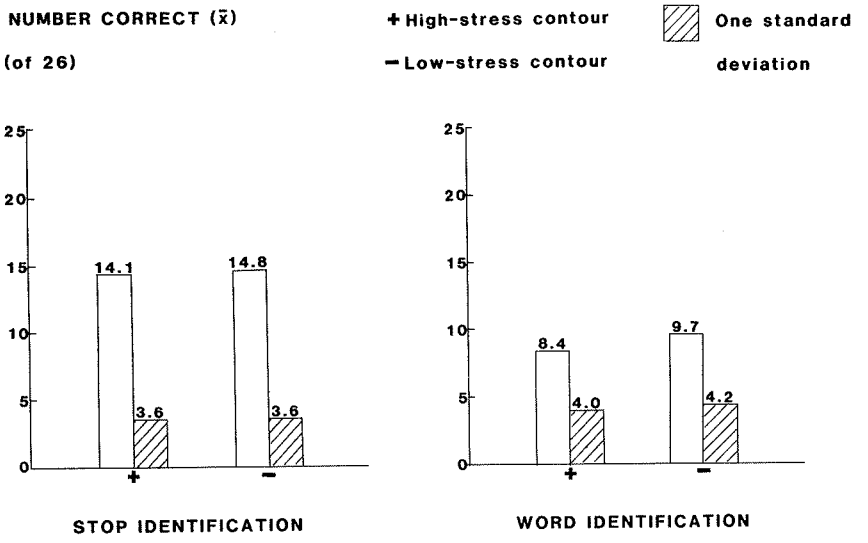

Figure 4. Mean number of correct identifications for stops and words in low-pass filtered sentences with high and low-stress fundamental frequency contours. (Group 2, hearing-impaired listeners.)

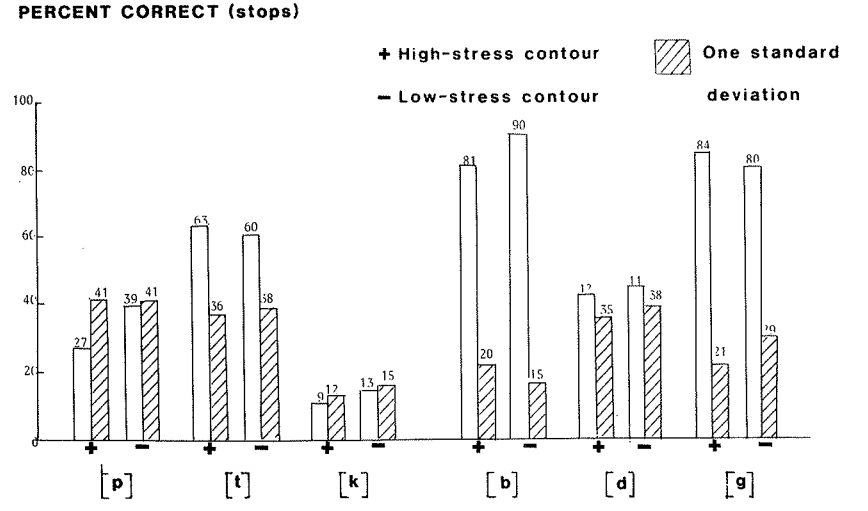

Figure 5. Correct initial phoneme identification in percent for Group 2 (Hearing-impaired).

presentation was particularly difficult or unusual.

In order to interpret these differences, subcategories of the normal-hearing group (Group 1) and the hearing-impaired group (Group 2) were made. The results were combined for low vs. high stress since those differences were generally not significant. The two groups were then divided up on the basis of best results (28 or more correct identifications for stops, Groups 1A and 2A) and worst results (Groups 1B and 2B). Figure 6 shows correct identification for these subcategories. The difference in stop responses between the best hearing-impaired group (2A) and the normal group as a whole (1) was highly significant $p<0.003$. The difference in word identification was also significant, $p<0.05$. This difference is, however, less convincing when the best hearing-impaired group is compared to the best normal group. The hearing-impaired group still performed better (34 correct vs. 30 correct for stops, 24 vs. 18 for words), but the differences were not significant, $p>0.05$.

Since any group can be subcategorized using a best-results criterion, the hearing-impaired group was divided into two new categories using pure-tone audiogram configuration criteria. The categories were (Group 2C) those ears most closely resembling the filter function used in the test, i.e. severity of hearing loss increasing sharply at a drop-off frequency lower than 1500Hz and at least a -35dB threshold drop between 1000Hz and 2000Hz and a threshold of less than -50dB(HL) at 2000Hz; and (Group 2D) those ears which least resembled the filter function, i.e. either the slope was too flat or the drop-off frequency was greater than 1000Hz (see Figures 7 and 8 for example audiograms). Correlation between the best-ear group (2A) and the most-like-filter group (2C) was high with all the ears occuring in 2C also being represented in 2A. Identification results for these two groups, for both stops and words, were also very similar, as can be seen in Figure 6. Differences in identification for both stops and words between

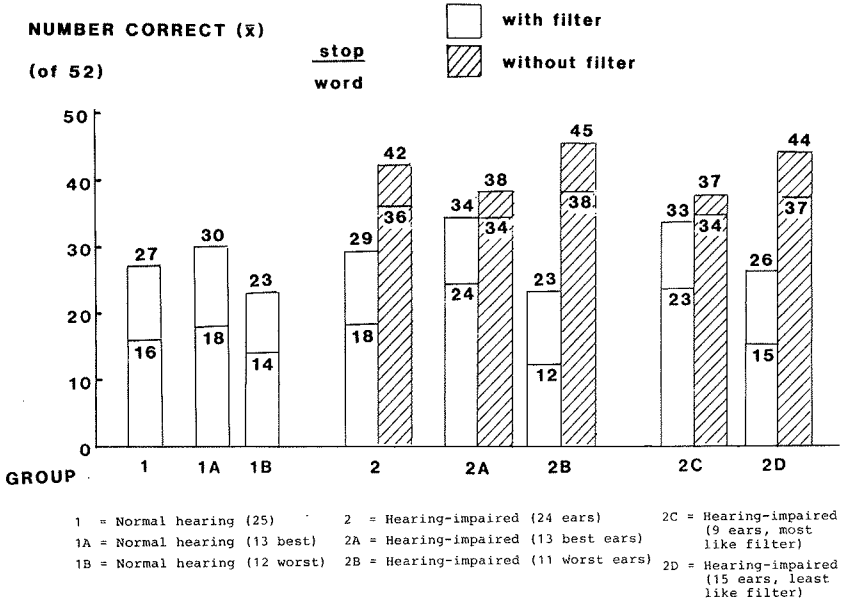

Figure 6. Mean number of correct identifications for stops and words for various group subcategories. (See text for subcategory criteria.)

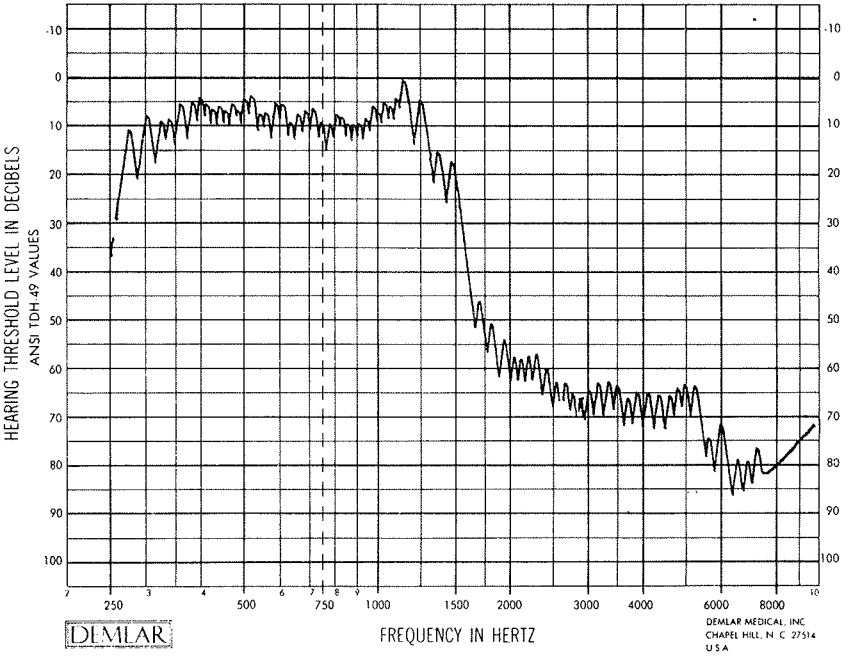

Figure 7. Example audiogram (Békésy) of a "most-like-filter" ear.

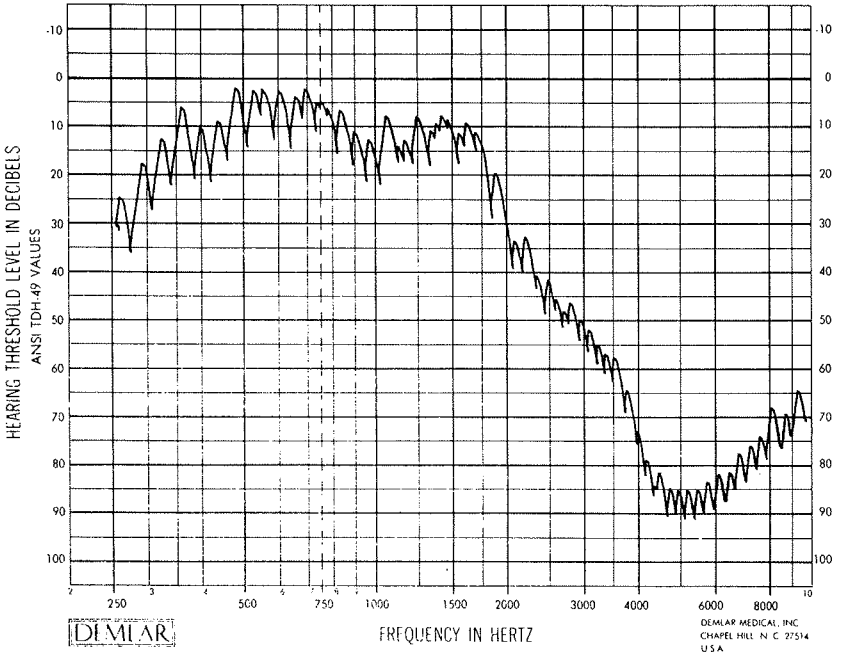

Figure 8. Example audiogram (Békésy) of a "least-like-filter" ear.

Group 2C (most-like-filter) and Group 1 (normals) were significant, $p<0.05$.

A final point of interest when comparing the hearing-impaired subgroups concerns listener reactions to the non-filtered test version. In general, those listeners who performed best on the filtered version noted little or no difference between the filtered and non-filtered versions, while those listeners who performed worst on the filtered version performed best on the non-filtered version (Fig. 6) and noted a considerable improvement in "clarity". There were, however, no significant differences which could be attributed to low or high stress. The normal listeners were not tested on the non-filtered version as an informal test indicated 100% correct identification.

4. DISCUSSION

A. Fo movement and stop identification

The results of these tests provide a negative answer to the question of whether or not increased movement of Fo in the vowels adjoining a stop can aid hearing-impaired listeners in identifying stops. The same results apply to normal listeners using filtered speech. In fact, in all groups and on all tests, identification of both stops and words was slightly better when the vowels carried low-stress pitch movement, i.e. lower Fo and little absolute movement. An explanation for this could lie in the fact that a lower fundamental produces a tighter series of harmonics which could in turn supply more energy to the critical formant amplitude shifts in the transitions thereby enabling better identification. The improvement, however, can only be seen as highly marginal as the differences were not statistically significant.

It seems then that at least regarding filtered speech in Swedish and for hearing-impaired listeners, heightened Fo

movement related to focus and sentence stress serves as a marker to direct the attention of the listener to the focussed word but does not intrinsically aid the perceptual mechanism in segmental resolution. It could be that the perception of Fo movement is integrated over a longer time interval than the perception of segments and, at least where global Fo movement is concerned, Fo movement perception is related to the pragmatic intentions of the speaker rather than to the phonemic content of the word in focus. If this is the case, then the perceptual mechanism could rely on increased intensity in the vowel and increased vowel duration during stress to aid in segmental resolution. During an informal listening session it was felt by several listeners that the high-stress stimuli sounded "thinner and weaker" than their low-stress counterparts. This could be due to the fact that the increased intensity normally associated with high stress was missing.

An aspect of production which could also contribute to a separation in perception of the two kinds of frequency movement dealt with here, i.e. Fo and formant frequency, is that of source. If perception of movement in the speech wave can be coupled to articulator movements as described by Fowler, et al. (1980), then the separate nature of the sources, i.e. tongue and jaw movement to alter resonance and laryngeal movement to alter fundamental frequency, could be perceived as relating to these separate sources from a production standpoint and therefore be processed separately. Clearly, these kinds of production-perception interactions and the processing of different kinds of movement need to be investigated further.

B. Compensation strategies and hearing-impaired listeners

Perhaps the most interesting result of the experiment pertains to the difference in performance between the normal-hearing listeners and the hearing-impaired listeners, especially those whose audiograms best matched the filter. The greatest difference in performance can be attributed to the tendency for hearing-impaired listeners to choose /t/ instead of /p/ or /k/

while the normal listeners tended to choose /p/. As the frequency of /t/-words dominated in the test material, the hearing-impaired group naturally came out ahead. A possible explanation for this could lie in the fact that listeners with a hearing loss, being accustomed to hearing speech resembling the filtered stimuli, were able to distinguish between the presence or absence of /p/ (low-frequency burst and low-frequency F2 transitions) while the low frequency nature of the filtered stimuli sounded labial to those unaccustomed to speech sounding similar to the stimuli. This would also account for the difference among members of the hearing-impaired group. Basically, the closer the correspondence between filter and hearing, the better able the listener is to comprehend the sentence.

While a hearing loss cannot be described as a filter function in absolute terms, the correlation in this experiment between performance results, filter frequency and audiometric configuration seems to indicate a certain performance predictability. Thus, if audiometric similarity is narrowly defined in terms of break frequency and slope, listeners tend to perform similarly when the filter function resembles the audiometric configuration. They even tend to perform better than listeners with normal hearing.

If we assume that those listeners with a hearing loss corresponding to the filter were able to discriminate between presence vs. absence of /p/, why then did they nearly always pick /t/ where the actual stimulus contained either /t/ or /k/? Were the better results simply a matter of chance, there being more /t/-words than /k/ or /p/-words in the test material? In a computer analysis of word frequency in Swedish newspaper material (Allén, 1972) the frequency of /t/-initial words used in the test was greater than /p,k/ in all but one case (Table 2). In the same material, t was also the most frequent letter of the six representing stops (Table 3). On the basis of this material, it could be tentatively conjectured that certain

	TAL 200	KAL	BAL	DAL 15	GAL
PAR 427	TAR 1964	KAR	BAR		
	TUR 137	KUR	BUR	DUR	
	TÅR 10	KÅR 12	BÅR		GÅR
PÅG	TÅG 55		BÅG		
	TAM			DAM 137	GAM
	TOK 8	KOK	BOK 876		

Table 2. Frequency of target words in a study of newspaper material (Allén, 1972).

	ABSOLUTE	RELATIVE
t	456035	8.238
d	225548	4.084
k	170580	3.089
g	166999	3.024
p	86131	1.560
b	64902	1.175

Table 3. Frequency of letters representing target-word initial stops used in the test. (From Allén, 1972)

hearing-impaired listeners, when presented with semantically non-redundant speech, choose the most frequent or probable word from the lexicon which matches the incomplete phonetic signal. It is possible that listeners can build up a frequency or probability strategy during the long period of time often associated with the progression of a noise-induced hearing loss. All listeners but one in the "most-like-filter" group were over 58 years of age, had very similar audiograms and performance, and had long histories of working in noisy environments. The one exception, 39 years of age, had of course a similar audiogram but did not perform nearly as well as the older listeners on the test.

An additional point of interest was that one listener with an asymmetrical loss performed better on the filtered stimuli test with his worse ear. This was noticed by the listener himself who expressed considerable surprise over it. His worse ear audiogram, however, fit the filter function almost exactly. His better ear loss began at around 2200Hz.

Finally the fact that persons with impaired hearing tended to choose words beginning with /t/ could have certain clinical implications. A greater awareness of compensation strategies both on the part of those using the strategies and those who are often in contact with persons suffering from a hearing loss could be instrumental in improving the communication ability of a relatively large group of people.

Further work with perception, hearing loss, and compensation strategies could also provide us with interesting insights into speech perception as a whole. It might be possible that the compensation strategies used constantly by the hearing impaired are also available to and used to a lesser degree by normal listeners when engaged in everyday speech in non-optimal, noisy environments.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Jarle Aursnes and the Department of Otorhinolaryngology, University Hospital, Lund, for making available both patients and equipment for the clinical part of the experiment. Thanks is also due to Bengt Mandersson for writing the pitch editing program, helping in constructing the stimuli and in interpreting the results of the experiment; and to Gösta Bruce and Eva Gårding for valuable discussion and comments.

REFERENCES

Allén, S. 1972. Tiotusen i topp, Almqvist & Wiksell, Stockholm

Cutler, A. 1976. Phoneme-monitoring reaction time as a function of preceding intonation contour. Perception & Psychophysics 20:1, 55-60

Fowler, C.A., Rubin, P., Remez, R., & Turvey, M. 1980. Implications for speech production of a general theory of action. In B. Butterworth (Ed.) Language Production, 373-420, London, Academic Press

Fry, D.B. 1958. Experiments in the perception of stress, Language and Speech 1:126-152

Gårding, E. 1967. Internal Juncture in Swedish, Lund, C.W.K. Gleerup

House, D. 1983. Perceptual Interaction between Fo Excursions And Spectral Cues, Working Papers 25:67-74, Department of Linguistics and Phonetics, Lund University

Jackendoff, R.S. 1972. Semantic Interpretation in Generative Grammar. Cambridge, Ma.: MIT Press

Johansson, G. 1975. Visual Motion Perception. Scientific American 232:6, 76-88

Johnson, T.L. & Strange, W. 1982. Perceptual constancy of vowels in rapid speech, Journal of the Acoustical Society of America 72: 1761-1770

Lehiste, I. 1970. Suprasegmentals. Cambridge, Ma.: MIT Press

Marslen-Wilson, W. & Tyler, L.K. 1980. The temporal structure of spoken language understanding. Cognition 8: 1-71

Pastore, R.E. 1981. Possible psychoacoustic factors in speech perception. In Eimas & Miller (Eds.) Perspectives on the Study of Speech, Hillsdale, N.J. Erlbaum

Picket, Revoile, and Danaher 1983. Speech-Cue Measures of Impaired Hearing. In Tobias and Schubert (Eds.) Hearing Research and Theory v.2, Academic Press

Risberg, A. & Agelfors, E. 1978. On the identification of intonation contours by hearing impaired listeners. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 2/3: 51-61

Van de Grift Turek, S. Dorman, M.F. Franks, J.R. & Summerfield, Q. 1980. Identification of synthetic /bdg/ by hearing-impaired listeners under monotic and dichotic formant presentation. Journal of the Acoustical Society of America 67: 1031-1039