# The encoding of fricatives for people with severe and profound hearing loss

*Deborah Anne Vickers*

Thesis submitted for PhD.

University College London

2003

***Dedicated to the Memory of***
***Alan Vickers***

## ABSTRACT

The aim of this work was to determine the potential for encoding voiceless fricatives for listeners with a severe-profound sensorineural hearing loss.

The first experiment determined whether normal hearing listeners could label band-pass filtered noises as fricatives. Flat spectrum noise bands varying in centre frequency, bandwidth and level were presented between two sinusoids to simulate a vowel-consonant-vowel syllable. Some combinations of noise centre frequency, bandwidth and level were consistently labelled as fricatives /f, s and ʃ/. Labelling was consistent with the acoustic features used to label natural fricatives.

The second and third experiments investigated the residual auditory abilities of profoundly hearing-impaired subjects. The second experiment determined the duration required to discriminate the centre frequency of bands of noise for several bandwidths, levels and centre frequencies. Most subjects could discriminate at least four, 250 Hz wide, spectrally adjacent noise bands within their residual hearing range at durations of 80ms or less, which is less than the duration expected for fricatives in running speech.

In the third experiment psychometric functions were measured for discrimination of the bandwidth of noise stimuli spectrally centred (linear scale) at 300 Hz. In task 1, either the wider (500 Hz) or the narrower (124 Hz) of two bandwidths was fixed, and the other was varied. All subjects were capable to some extent of discriminating bandwidth. In task 2, either the upper or lower frequency difference between the stimuli was eliminated. Subjects tended to use high-frequency edge cues to discriminate the noises when the wider noise band was fixed. When the narrower noise band was fixed, the noise edge used varied across subjects. Together the second and third experiments indicated a potential for using noise bands to encode fricatives within the low-frequency hearing range of hearing-impaired listeners.

The final experiment investigated the ability of hearing-impaired listeners to label noise bands as fricatives when presented within their residual hearing range. The stimuli were encoded VCV syllables, where the vowel was represented by a sinusoid and the consonant was an encoded version of either /f, s, ʃ/. All subjects were able to some extent to label the encoded stimuli as fricatives and most subjects performed best when listening to noises with the spectral envelope simplified to one spectral peak and lowered by a factor

of 8 relative to the natural token.


## ACKNOWLEDGEMENTS

I am indebted to Andy Faulkner for his supervision. He provided useful and insightful comments to enrich the work. His interest in the work has always been a great motivation.

The other members of the department of phonetics at the time when I worked there provided a supportive team in which ideas could be discussed and thought through. The members of the team were Stuart Rosen, Richard Baker, Jianing Wei, Evelyn Abberton, Adrian Fourcin and Valerie Hazan. Jianing Wei and Richard Baker were of particular help with computer programming and Stuart Rosen was a source of inspiration for ideas.

My friends in the experimental psychology, hearing lab in Cambridge helped me to keep my sanity and supported me while working on this thesis part-time. I must offer particular thanks to Brian Moore for his wonderful ideas and comments, Brian Glasberg for in depth discussions about statistics and Tom Baer for his help with Matlab. Michael Stone, Marina Rose, Martina Huss, Geoff Moore, Andrew Oxenham, Josephine Marriage and Jose Alcantara were all soul mates that helped me greatly throughout this work.

Finally I must thank my family, my husband John who gave me an incredible amount of support and encouragement and our two young daughters, Emma and Andrea who helped to keep everything in perspective. I am also grateful to my parents Janet and Alan Vickers and my sister Diane and brother Steven have always been very positive and supportive.

# CHAPTER ONE

## 1 INTRODUCTION

It is well known that listeners with severe-to-profound hearing impairments have problems understanding speech in quiet and in the presence of background noise. There are many interacting factors influencing the speech perception abilities of these listeners, such as: a low level of audibility, impaired spectral and temporal processing, and loudness recruitment. In addition to this, the limited audible frequency range that is often observed in severe to profoundly hearing-impaired listeners may effectively remove crucial speech perceptual cues. Typically the cues lost are the weak, high frequency, and aperiodic sounds present during voiceless excitation, which are useful cues for the identification of consonants.

The main aim of this thesis is to determine if some of these speech perceptual problems can be overcome by encoding the high-frequency energy present in voiceless fricatives and presenting it within the residual hearing range of the listener.

To achieve this aim, experiments were conducted to look at: whether, for normal-hearing listeners simple noise bands could carry relevant information for labelling voiceless fricatives; the psychoacoustic abilities of listeners with severe-to-profound hearing impairments for discriminating bands of noise within their residual hearing range; the effectiveness of a range of processing strategies for encoding voiceless fricative information within this range.

This introduction describes the prevalence of hearing impairment and contains a discussion of the psycho-acoustical abilities of and speech perceptual difficulties encountered by hearing-impaired listeners. There is also an account of some of the currently available approaches for improving speech perception for hearing-impaired listeners.

### 1.1 The prevalence of hearing impairment

The results from the National Study of Hearing, summarized by Davis (1995), showed that in the United Kingdom there are approximately eight and a half million people with a four-frequency average (4FA) hearing loss (averaged over 500, 1000, 2000 and 4000 Hz) greater than 25 dB (approximately 20 % of the adult population). There are approximately 170,000 adults with profound hearing impairments (4FA hearing loss greater than 85 dB) and 470, 000 adults with severe losses (4FA greater than 60 dB). Hearing impairment is mainly seen for older people, with 90 % of people with a 4FA hearing loss greater than 65 dB being older than 50 years. For the population aged 60 and older, approximately 1% has a 4FA hearing loss greater than 85 dB (approximately 140,000 people) and 3% have 4FA losses greater than 65 dB (approximately 420,000 people). For the purpose of this thesis, it is worth noting the prevalence of high-frequency hearing impairment. There are approximately 3.5 million adults with high-frequency losses (averaged over 4, 6, and 8 kHz) greater than 65 dB (8.7% of adult population), again 90% of who are over 50 years old.

## 1.2 Psychoacoustic abilities of severely and profoundly hearing-impaired iisteners

The commonest cause of hearing loss resulting in a severe-to-profound impairment is that of cochlear damage. Cochlear damage is associated with a reduced ability to understand speech and in particular with difficulties in the presence of background noise. The loss of audibility is one factor that limits speech perceptual abilities in people with severe and profound cochlear hearing loss, but other suprathreshold factors such as abnormal growth of loudness and reduced spectral and temporal analysis are also likely to play a large role.

### *1.2.1 Loudness recruitment*

Most people with a hearing loss due to cochlear damage exhibit 'loudness recruitment' (Fowler, 1936). This refers to abnormal loudness perception. Hearing-impaired listeners have a higher than normal threshold for detecting a sound, but once the level of that sound is increased beyond threshold, the growth of loudness with increasing sound intensity is more rapid than normal for levels up to about 90-100 dB SPL. At 90-100 dB and higher the sound is perceived by a hearing-impaired listener to have roughly the same loudness as that perceived by a normal-hearing listener. This abnormal growth in loudness leads to a reduced dynamic range. Moore et al. (1997) tested a group of unilaterally hearing-impaired listeners on a matching task where the stimuli were sinusoids, amplitude modulated at rates of 4, 8, 16 and 32 Hz (to cover some of the most important modulation rates observed in the temporal envelope of speech). Subjects were required to adjust the modulation depth of the sinusoid in one ear to match the perceived amount of fluctuation of the modulated sinusoid in the other ear. They found that subjects matched a given modulation depth in the impaired ear with a larger depth in the normal ear. This showed that loudness recruitment had the effect of exaggerating the perceived modulation depth in the impaired ear. This has implications for the perception of speech, music and environmental sounds because the dynamic fluctuations within those stimuli would be exaggerated, making the softer parts inaudible and the more intense parts would have a loudness level near that for a normal-hearing listener.

Listeners with a limited dynamic range require a hearing aid that contains some form of gain limiter, or, preferably some form of compression system to avoid sounds becoming painfully loud. It is often the case that listeners with severe-to-profound hearing-impairments require a large amount of compression to avoid loudness discomfort. This has the effect of smoothing the temporal envelope cues, which potentially could be very important to severe-to-profoundly hearing-impaired listeners for the perception of speech.

### *1.2.2 Frequency Selectivity*

Pick et al. (1977) used a comb-filtered noise technique (Houtgast, 1972) for estimating frequency resolution and auditory filter bandwidth in normal and hearing-impaired listeners. There was a large degree of scatter in the data but the general pattern of results showed that, for listeners with mild losses, the auditory filter bandwidth typically fell within the normal range. As absolute threshold increased, the bandwidths of the auditory filters increased. For listeners with thresholds of 40-50 dB HL, the bandwidth was approximately twice that of normal listeners and for listeners with thresholds of 70-80 dB HL, the bandwidth was approximately three times normal.

Faulkner et al. (1990) measured psychophysical tuning curves for signal frequencies of 125 and 250 Hz in a group of subjects with profound hearing losses. They estimated auditory filter bandwidths to be 2 to 3 times greater than normal. Glasberg and Moore (1986) tested a group of unilateral and bilaterally hearing-impaired listeners with moderate to severe losses. They measured auditory filter shapes using the notched-noise method (Patterson and Nimmo-Smith, 1980) for centre frequencies of 0.5, 1.0 and 2.0 kHz. All of the subjects with severe impairments showed filters 2 to 3 times broader than normal. Some of the effect of broadening occurs just because of the higher sensation levels that the hearing-impaired listeners are tested at. For normal hearing listeners as the level of the probe increases the filters become broader but it still remains a fact that even when comparing the same intensity level of the probe that the measured auditory filters are much broader for hearing-impaired listeners. This considerable loss of frequency selectivity causes difficulties in resolving the spectral structure of speech, as the internal representation of the signal is "smeared" compared to normal. The smearing of the internal representation is particularly disabling for sounds that sometimes contain closely spaced peaks in the spectral envelope, which can be important cues for identification e.g. vowels, semi-vowels, and nasals. Fewer frequency resolution channels can also lead to difficulties in perceiving speech in the presence of background noise or a competing speaker. This effect is highlighted in studies using cochlear implant simulations to observe the effect of channel number on speech perception in noise (Dorman et al., 1998; Fu et al., 1998). These studies show that performance on speech perceptual tasks in the presence of noise worsens as the number of channels decreases, showing that spectral resolution plays an important role in picking out target speech from a background noise.

### *1.2.3 Temporal Resolution*

One of the problems associated with measuring temporal resolution in hearing-impaired listeners and comparing the results to those obtained from normal hearing subjects is that testing is often carried out at lower sensation levels (SL's) for the hearing-impaired listeners. This is because hearing-impaired listeners suffering with loudness recruitment cannot listen at high sensation levels. It has been shown that normal hearing listeners perform less well on temporal resolution tasks at low sensation levels than at higher sensation levels (Plomp, 1964; Shailer and Moore, 1983). On many measures of temporal resolution, if hearing-impaired listeners are tested at the same sensation level as normal hearing listeners, their performance is very similar to that of normal listeners. When tested at the same sensation level there are some tasks in which hearing-impaired listeners perform slightly less well than normals. This is often the case for tasks using noise stimuli, e.g., in the detection of temporal gaps in bandlimited noise, (Glasberg et al., 1987). This can be attributed to loudness recruitment that exaggerates the inherent fluctuations in the noise thus making it more difficult to perform the specific temporal resolution task. For deterministic stimuli containing no inherent fluctuations, hearing-impaired listeners tested at the same sensation level sometimes perform slightly better than normals, e.g., in the detection of temporal gaps in sinusoids (Moore and Glasberg, 1988; Moore et al., 1989).

It is important to note, however, that hearing-impaired listeners with a severe or profound loss typically have to listen at very low sensation levels due to their high audibility thresholds. Therefore, in practice, listeners with severe-to-profound losses would be expected to have poorer than normal temporal resolution abilities. Despite this these listeners do usually retain some level of temporal processing.

## 1.3 Speech perceptual abilities

The majority of hearing-impaired listeners have the greatest hearing loss in the high frequencies, where speech cues are typically of low intensity. Audibility cannot completely account for the poor performance of these listeners in speech perceptual tasks, and other suprathreshold factors must play a role. Some of the suprathreshold factors have been discussed above, along with the effect they have on speech perception, but there are a few other important issues regarding speech perception not discussed so far. These are: the effect of a reduced frequency range on the use of temporal cues in speech perception and the role of auditory-visual integration.

### *1.3.1 The effect of a reduced frequency range.*

Miller and Nicely (1955) carried out a consonant identification task with normal-hearing listeners to examine the effect of masking noise and filtering on the perception of consonants. They determined crossover points on the frequency axis for which the region below the crossover point had equal importance for detection of the feature being analysed as the frequency region above the crossover point. They found that the crossover point for most features was towards the lower frequency end of the frequency axis indicating that the lower frequencies carry the most weight for perception of the majority of features. The crossover points were: 450 Hz for nasality; 500 Hz for voicing; 750 Hz for affrication; 1900 Hz for place of articulation; and 2200 Hz for the duration of noise component in s, ʃ z and zh as opposed to other consonants. This suggests that hearing-impaired listeners with good low-frequency hearing should be able to utilize cues for voicing and nasality and perhaps affrication, but place of articulation and duration of the aperiodic sibilant energy could be difficult to detect.

Bilger and Wang (1976) tested 22 hearing-impaired listeners, with a range of audiometric configurations on a consonant perception task. They found that audiometric configurations gave a good indication of the pattern of results. They divided listeners into three categories: normal or mild flat losses, moderate to severe flat losses, and high-frequency losses. The high-frequency loss group could not perceive the sibilance feature (high-intensity frication) at all but did relatively well with voicing and nasality. The results for listeners with flat moderate to severe losses showed that they were typically able to identify sibilance, duration and voicing but had some difficulty with place of articulation and struggled to detect the nasality feature. The listeners with normal or mild flat losses perceived most features reasonably well, but with some place of articulation errors and voicing and frication not perceived as well as the other features. Extrapolating from these results, it would appear that a listener with a moderate to severe loss in the low frequencies and a profound or total loss in the high frequencies would have difficulty with contrasts of place of articulation and sibilance, but may be able to perceive voicing, nasality and duration.

Walden et al. (1981) carried out a study with unilaterally hearing-impaired subjects using filtering to match the audiometric configuration of the normal ear to that of the impaired ear. They found large differences in speech perceptual performance between the impaired ears and the normal ears with the audible spectrum matched by filtering to the audible spectrum for the impaired ear. They only found similarities when the impaired ear had normal thresholds at low frequencies with a rapid transition to a high-frequency hearing loss. The results of Walden et al. (1981) showed that if there is a hearing loss in the high frequencies with relatively good hearing in the low frequencies, the performance of the subjects could be modelled by simply filtering speech to match the cut-off frequency of the loss. However, many listeners with high-frequency hearing losses do not have near normal hearing in the low frequencies, so a simple filtering approach is not sufficient to model their performance.

Fabry and van Tasell (1986) found that threshold elevation and filtering was sufficient for simulating hearing impairment in the normal ear of 4 out of 6 subjects with high-frequency losses in their impaired ears. They suggested that for the other two test subjects, other factors such as impaired temporal and/ or frequency discrimination would have caused the difference in performance.

Zeng and Turner (1990) investigated whether, by making the low frequency or high-frequency energy in CV (consonant-vowel) syllables audible to hearing-impaired listeners, they could perform at the same level as normal hearing listeners. They tested three hearing-impaired subjects, two with severe losses and one with a moderate loss. They found in a CV fricative (/s, ʃ f, θ/) identification task that when only the formant transition region was made audible, performance was not as good as for normal-hearing subjects, given the same degree of audibility. However, if the high-frequency regions containing cues to the spectral shape of frication noise were made audible, then their subjects performed similarly to a group of normal subjects at the same level of audibility. This finding suggests that many hearing-impaired listeners retain suitable residual abilities to discriminate the gross high-frequency spectral differences associated with the obstruents, while they are less able than normals to utilize formant transition cues.

The results of the studies described above suggest that the removal of the higher frequencies (above 750 -1000 Hz), either by lowpass filtering for normal-hearing listeners or as a consequence of hearing loss for impaired listeners, has detrimental effects on the perception of fricatives, affricates and plosives. Some caution should be used in interpreting situations where the normal-hearing listeners perform better than the hearing-impaired listeners for a similar low-pass condition. This is because normal-hearing listeners have been shown to utilize the information from the transition band (i.e. region between edge of pass-band and complete cutoff) effectively even when the cutoff slope is 100 dB/octave (Warren and Bashford, 1999). This may account for differences observed when the hearing-impaired listeners have good low-frequency hearing, but is unlikely to account for the vast discrepancies in scores when the hearing-impaired listeners have poor hearing in the low frequencies.

If hearing is well preserved at low frequencies, then vowel formant-frequencies (F1 and possibly F2) and their transitions, voicing energy and nasality can be perceived. In people with this type of hearing loss, amplification to make the high-frequency speech cues audible might be sufficient to give near-normal speech perception performance, as long as there are not any dead regions (regions without functioning inner hair cells and/or neurones; (Moore et al. 2001)) in the high frequencies. If, however, hearing is impaired in the low-frequency region as well then there is little redundancy amongst the acoustic cues available to the hearing-impaired listener and it may be difficult to use formant transitions. For these listeners, it is likely that suprathreshold factors play a large role in their speech perceptual performance and that signal processing aids, which overcome some of the deficits in suprathreshold processing, would be beneficial. For listeners with severe-to-profound hearing impairment, it has been shown that frequency selectivity is typically two to three times worse than normal while temporal resolution is relatively well preserved (Rosen et al, 1990). It is therefore likely that these listeners will rely substantially on temporal cues and relatively gross spectral cues for speech perception.

### *1.3.2 The role of temporal cues in speech perception*

Work looking at the importance of temporal resolution for speech perception has highlighted the important role that temporal cues can play, especially for the hearing-impaired listener, for whom spectral resolution may be dramatically reduced (van Tasell et al., 1987; Rosen, 1992; Shannon et al., 1995).

Rosen (1992) described three main categories of temporal cues available for identifying speech, distinguished by frequency range.

1. Envelope cues (<50 Hz) refer to fluctuations in the overall envelope.
2. Periodicity refers to the presence or absence of voice periodicity (50 Hz to a few hundred Hz), which provides segmental information regarding voicing and some manner information, and prosodic information regarding intonation and stress.
3. Fine structure (600 Hz and above) refers to variations in wave shape in a single period of a periodic tone, or over short intervals of an aperiodic sound. Fine structure contains information relating to formant structure.

The importance of each category of temporal cue is probably dependent to some extent on the degree of residual frequency selectivity. It is likely that envelope and periodicity cues play a large part in the speech perceptual abilities of severe-to-profoundly impaired listeners.

Van Tasell et al. (1987) suggested that envelope cues in conjunction with visual information should give near-normal performance for speech perception with the use of training. The aim was to look at the effect of envelope bandwidth on consonant perception. The stimuli used in the experiment were pink noise carriers multiplied by the envelope of a specific VCV (vowel-consonant-vowel) utterance. There were 3 different cutoff frequencies for the speech envelope, at 20, 200 and 2000 Hz. 19 consonants were used. Each was played four times in a random order within a block and only one noise condition was used for each block. There was also an unprocessed condition. The normal hearing subjects listened to the stimuli and stated which consonant they had perceived. The results of multi-dimensional scaling showed that the envelope bandwidths could be used to define envelope-based consonant classes, which were termed envemes. Four classes were derived having similar envelope characteristics within each class. The classes were: (1) voiceless stops; (2) voiced stops and voiced fricatives; (3) voiceless fricatives; and (4) nasals and approximants. Van Tasell et al. proposed that the consonants falling within each enveme class could be discriminated with the use of lipreading cues. Therefore if information regarding envelope cues could be given to the hearing-impaired listener in conjunction with lipreading they should perform extremely well.

Faulkner and Rosen (1999) studied the role that temporal cues could play in auditory-visual speech perception using a range of spectrally invariant signals. The auditory signals used were: a pulse train of fixed frequency and amplitude, present only during voicing; a pulse train of fixed frequency during voicing with additional amplitude envelope variations related to original speech envelope changes; a pulse train of fixed amplitude during voicing that varied in frequency according to the fundamental frequency of the voice; a pulse train during voicing varying in both frequency and amplitude. They found that the pulse train that was gated on to indicate voicing had considerable effect, while only a small increase in performance on a speech perceptual task was obtained with the addition of information to cue fundamental frequency and amplitude envelope changes.

Faulkner and Rosen (1999) also studied the effect of using noise to cue the aperiodic energy present in voiceless obstruents. They used two processing conditions, both involving the pulse train varying in frequency and amplitude. In the first condition, the pulse train was combined with a fixed-level noise present during voiceless excitation. In the second condition, the noise varied in amplitude in accordance with the envelope of the original speech. They found that indicating the presence of voiceless excitation had a significant effect on consonant perception, due to improved voicing and manner detection. There was little extra benefit gained from the cues to the amplitude fluctuations of voiceless speech.

The results of these studies show that temporal encoding of speech, particularly when the auditory information is used in conjunction with lipreading can convey considerable information. With the use of lipreading, minimal temporal information is required – e.g., gross indications of the presence of periodic and aperiodic speech, with details of amplitude envelope mattering little.

## 1.4 Auditory-visual speech perception

Many hearing-impaired listeners rely heavily on lip reading to compensate for their reduced ability to extract information from the auditory signal. Visual information is also important to normal hearing listeners in normal speech communication. The importance of visual information is shown by the 'McGurk' effect (McGurk and MacDonald, 1976). To demonstrate the McGurk effect, listeners were presented with conflicting auditory and visual information in a phoneme perception task, for example a visual signal of the utterance /ka/ and an auditory signal of the utterance /ba/. The resulting percepts were often of a phoneme that was different from either of the ones uttered. This indicates the integration of information from the auditory and visual domains. This shows that subjects were not mostly reliant on one modality but use information from both modalities to perceive speech.

Auditory-visual integration has important implications for speech perception in hearing-impaired listeners, because, fortunately, the information that is typically lost due to hearing impairment is relatively easy to extract via lipreading (Summerfield, 1987). Auditory information cueing place of articulation is mainly found above 1900 Hz (Miller and Nicely, 1955). This information is inaudible for listeners with high- frequency hearing losses. Background noise also affects the perception of this information because the acoustic cues have very low levels and are easily masked. It is therefore not surprising that the majority of hearing-impaired listeners are heavily reliant on lipreading. Another feature is that the low-frequency cues for voicing and manner (amplitude envelope variations, temporal gaps and nasality), which are difficult to perceive through lipreading, are fairly robust auditory cues and usually are available to hearing-impaired listeners to some extent. Grant et al. (1998) supported this idea by showing that a group of hearing-impaired listeners processed auditory-visual utterances by extracting the place information visually and the manner and voicing auditorily.

It might be expected that lipreading skills would in some way be dependent on the degree of hearing loss or other audiometric measures, many believe that as hearing loss becomes greater lipreading skills improve. Bernstein et al. (2000) showed that there is a relationship between lipreading skills and exposure to natural speech. They compared visual speech perceptual performance with a range of audiological measures, and found that the only significant correlation was a negative one with "years since last used hearing aid". Another interesting correlation found by Berstein et al. was that hearing-impaired teenagers who had been brought up with an oral background scored more highly on a lipreading test than those who had been brought up with manual communication.

These results suggest that lipreading performance in adult life is an ability that is dependent on an individual's pre-disposition to be a good lipreader and on previous use of an auditory signal in the development of auditory-visual speech communication abilities.

For those with profound high-frequency loss, conventional amplification goes some way towards providing auditory support for lip reading by transmitting the low- frequency temporal information relating to voicing, intonation and amplitude envelope, which can be combined with the place cues obtained via lipreading, thus maintaining good auditory-visual speech perception abilities.

This, of course, does not mean an auditory signal that provides cues to manner and voicing is sufficient for allowing excellent auditory-visual speech perception. Normal-hearing listeners can utilize an abundance of acoustic cues, making the perception of speech relatively effortless. In situations where the auditory signal is drastically degraded, for example by hearing impairment or the presence of background noise, speech perception requires a great deal of concentration and a reliance on face-to-face communication. Normal-hearing listeners try to compensate for cues lost in the presence of background noise by utilizing the remaining cues available to them in addition to lipreading. For severely and profoundly hearing-impaired listeners there is seldom a redundancy amongst cues, so hearing aids need to go some way to lightening the perceptual load and providing some place of articulation cues.

Conventional amplification does not typically provide severe-to-profoundly impaired listeners with high-frequency information. It is therefore necessary to look at different approaches for presenting the place cues and other high-frequency cues, such as voiceless excitation, that are typically found above 1900 Hz. The approaches considered here are: frequency transposition, speech pattern extraction, and cochlear implants.

## 1.5 The effectiveness of hearing aids

Conventional hearing aids are primarily designed to restore audibility of speech for the hearing-impaired listener. This typically works well for listeners with mild and moderate cochlear hearing losses (Pavlovic, 1984; Kamm et al 1985), but for people with severe or profound hearing impairments the restoration of audibility is often impossible, and when it can be achieved, is not sufficient to give good speech discrimination.

Flynn et al. (1998) studied the speech perceptual performance of two groups of listeners wearing their own well-fitted linear conventional hearing aids. One group had severe hearing-impairment (3FA of 0.5, 1.0, & 2.0 kHz = 61-80 dB HL) and the other severe-to-profound (3FA of0.5, 1.0, & 2.0 kHz = 81-100 dB HL) hearing impairment. Performance was assessed both in quiet and in a multi-talker babble with a signal-to-noise ratio (SNR) of +10 dB. Vowel tests, consonant tests, monosyllabic word tests and sentence tests were conducted in quiet and the sentence speech material was also used for testing in noise. Both groups of subjects produced high scores in vowel perception tests. The average score was 94% for the severely hearing-impaired group and 84% for the severe-to-profoundly impaired group. The mean scores for the consonant test were 73% and 51% for the severe and severe-to-profound groups respectively. The severely hearing impaired scored 67% on average for monosyllables whereas the severe-to-profound group only scored 39%. For the sentence test, the mean scores for the severely impaired were 83% in quiet and 74% in noise. For the severe-to-profoundly hearing-impaired group, the mean sentence score in quiet was 56% while in noise it was 34%. Large individual differences were apparent in the results. These results highlight how poorly listeners with severe-to-profound hearing impairments perform when using traditional linear amplification, even when contextual information is available.

### 1.5.1 Compression hearing aids

Loudness recruitment can often cause discomfort so a form of gain limiting is typically prescribed to avoid discomfort from loud sounds and to reduce the intensity range of speech to fit more completely into the audible dynamic range. The most common form of gain limiting is that of compression. Compression has the effect of improving the audibility of low level speech and maintaining good speech perception abilities over a wide range of sound levels (Moore and Glasberg, 1986).

There are a wide range of approaches to providing amplitude compression within hearing aids, this discussion will focus primarily on two types of compression categorized by the role they have as: as compression limiters; or as syllabic compressors. The compression circuits vary in the rate with which the gain decreases in response to increases in the input signal (determined by the attack time) and also on the time taken for the gain to recover when the input level decreases (determined by the release time, also called the recovery time).

To avoid sound levels becoming uncomfortably loud, compression limiters respond when sound levels become too high (approximately 85 dB SPL). These circuits typically have a short attack time (1-10 ms), a recovery time between 20-100 ms, and a high compression ratio (10). A different purpose of compression is to restore loudness perception to that of normal hearing listeners. Compressors of this type are called syllabic compressors, because the amount of compression is rapidly changing. The compressors have a threshold of about 45 dB SPL, have an attack time between 1 and 10 ms, a recovery time between 20 and 150 ms, and a compression ratio up to 3.

These two forms of compression have different effects on the temporal envelope of speech. The compression limiter only responds at high levels of the input signal but once triggered, smoothes out many of the peaks in the temporal envelope due to the high compression ratio. Syllabic compressors typically have low compression ratios and therefore the overall shape of the envelope is retained. The other factors affecting the temporal envelope of the output are the attack and release times. If these are short they can impose spurious transients upon the temporal envelope, but if long they can leave the overall envelope shape relatively untouched. The long release times are used to reduce waveform distortion. Typically fast attack times and slow release times allow for a compromise between the necessity to limit the high level peaks and to avoid too much distortion of the envelope cues.

Compression hearing aids work well for listeners with mild-to-moderate losses; such losses typically do not require high compression ratios. The work of Souza and Turner (1999) supported the claim that conventional linear aids work well for listeners with mild to moderate losses. They compared the performance of hearing-impaired listeners using linear and compression amplification and found that, at low levels, the compression amplification gave greater audibility of the input signal and thus better speech discrimination. As the input level increased to mid levels, audibility became similar for linear and compression amplification, and speech perceptual scores correspondingly became similar. They concluded that, for listeners with mild to moderate losses, hearing aids that could give high levels of audibility were sufficient to give good speech perceptual performance. However, for listeners with severe or profound losses it would not be possible to give sufficient audibility; even with compression hearing aids, performance would fall a long way below normal and the compression ratios required would result in severe distortion of the temporal envelope. Drullman and Smoorenburg (1997) tested a group of 16 profoundly hearing impaired listeners with a six-channel instantaneous compression system. They looked at compression ratios of 1, 2, 3 or 5. The subjects were tested with audio-visual sentence test materials and syllable intelligibility was measured. All subjects performed better in the audio-visual condition than in visual alone showing that they were able to utilize information from the auditory signal to help with speech perception. They didn't find a significant improvement for compression compared to linear amplification and they found that high compression ratios caused a decrease in performance. Only one of the sixteen subjects appeared to gain benefit from the use of compression. This suggests that profoundly hearing-impaired listeners are able to use the temporal information in speech to help with audio-visual speech perception and that they are sensitive to degradation of the temporal envelope by compression.

### *1.5.2 Difficulties associated with high-frequency amplification*

The majority of people with severe or profound hearing impairments tend to have the greatest losses at high frequencies, where speech cues are typically of low intensity. There have been several studies looking at the benefits of high-frequency amplification in listeners with high-frequency hearing impairments (Murray and Byrne, 1986; Hogan and Turner, 1998; Turner and Cummings, 1999). The results showed that, for some listeners, performance improved with an increase in the audible high-frequency range. However, the performance of other listeners did not change and there were some listeners whose speech perception abilities actually deteriorated. Turner (1999) summarized the pattern of results for listeners with high-frequency hearing losses by suggesting that if the hearing loss at a particular frequency was less than 40 dB, then the addition of high-frequency speech information was beneficial. However if the hearing loss was greater than 60 dB above 2.5 to 3 kHz, then the addition of extra high-frequency energy was typically of no benefit. Moore and Glasberg (1997) attributed the deterioration of performance with the addition of high frequency information to the presence of "dead regions". Dead regions are sections of the cochlea where there are no functioning inner hair cells and/or neurones (Borg et al., 1995). Moore and Glasberg proposed that, if a listener has a dead region, then as the stimulating frequency falls within the frequency range of the dead region, speech perception might deteriorate.

Dead regions can be measured in subjects with hearing impairments. Moore et al., 2000 described the use of a test called the TEN test. The TEN test involves detecting sinusoids in the presence of a broadband noise designed to produce equal masking across frequency, the noise is called the threshold equalising noise (TEN). The detection threshold is approximately equal to the level of the noise in a one-ERB (equivalent rectangular bandwidth) wide band centred at 1 kHz. For example for normal hearing a noise level of 70 dB/ERB usually elicits a masked threshold of 70 dB SPL. An abnormally high masked threshold at a specific frequency indicates the presence of a dead region. Vickers et al. (2000) looked at consonant perception for a group of subjects with and without dead regions measured using the TEN test. Subjects listened to VCV (intervocalic consonant) stimuli over HD580 headphones with amplification presented in accordance with the "Cambridge" formula (Moore and Glasberg, 1998). The speech was low-pass filtered at a range of cut-off frequencies. They found that for subjects without dead regions performance increased as the cut-off frequency increased showing that they benefited from high frequency amplification. For subjects with dead regions performance increased with increasing frequency to just above the cutoff frequency of the dead region. For cutoff frequencies above the dead region cutoff frequency there was either no change in performance or deterioration in performance with increasing cutoff frequency.

The concept of dead regions has important implications for conventional amplification, because it implies that it is more appropriate to limit the audible frequency range than to extend it when there is a severe or profound loss at high frequencies (usually indicative of a dead region). Individual differences in the presence/absence and extent of dead regions could provide an explanation for the wide range in performance with increasing hearing loss that was described by Flynn et al. (1998). Interestingly the work of Skinner and Miller (1983) looking at the relationship between amplification bandwidth and speech intelligibility found that the speech perception scores for a group of listeners who most likely did not have dead regions went down as the bandwidth was reduced.

The studies described above go some way towards explaining why conventional amplification falls short of providing severe-to-profoundly hearing-impaired listeners with sufficient information for good speech reception. The problems may arise due to: hearing aids not being capable of providing the gain required to give sufficient audibility, distortion of the temporal envelope due to high compression ratios and rapid response times, or the presence of dead regions.

For many subjects with a severe or profound hearing impairment, the useful audible frequency range may be very small, and in addition, other suprathreshold factors may impair speech discrimination. It is therefore important to investigate possibilities other than simple amplification for improving speech perception in these listeners.

## 1.6 Approaches to improving speech perception for the severe-to-profoundly hearing impaired

### *1.6.1 Frequency Transposition*

Psychoacoustic studies have shown that the majority of severe-to-profoundly hearing-impaired people retain potentially useful auditory abilities which are not fully exploited by conventional amplification, and which could be utilised to increase the speech information perceived by these listeners. Amongst these residual abilities are: 1) Significant frequency selectivity in frequency regions of less profound hearing loss (Faulkner et al., 1990); 2) The ability to distinguish between periodic and aperiodic stimuli with a similar frequency content (Rosen et al., 1990); 3) Relatively good resolution of amplitude modulations of low-frequency carrier signals at modulation rates of 80 Hz and below (Faulkner and Rosen, 1990).

In combination, these three residual abilities suggest that there should be considerable potential for the use of frequency transposition techniques intended to shift the temporal and spectral information carried by higher-frequency speech cues to the lower-frequency receptive ranges of these listeners. There have been many attempts to explore different techniques by which parts of the speech spectrum can be moved into a lower frequency range (e.g. Johannson, 1966; Velmans, 1973). However, none of these attempts has produced more than marginal success (Braida et al, 1979). There are at least two possible explanations for the poor results. Firstly, transposition has the effect of increasing the complexity of the signal presented to the listener, especially in terms of spectral structure. Since these listeners have impaired analytic abilities, such an increase in complexity may well act to further reduce the match between the acoustic information and the listener's receptive abilities. Secondly, transposition processing did not preserve crucial features of the spectral shape of the original high-frequency part of the speech.

A more recent development in frequency transposition is the AVR Transonic frequency transposing hearing aid (Rosenhouse, 1990). Here, when the high frequency region (above 2.5 kHz) is dominant, the spectrum of the energy above 2.5 kHz is compressed by a chosen factor to move the high-frequency energy into the low-frequency region, maintaining the relative relationship of the spectral peaks. McDermott et al. (1999) tested five adults with the Transonic hearing aid and showed that four of the five subjects performed better with the Transonic than with their conventional aid. For two of the subjects, a large proportion of the benefit could be attributed to the good low-frequency amplification provided by the Transonic. These two subjects did not show an improvement in the perception of cues associated with frequency transposition. Also the scores were similar for all settings on the AVR, one of which presented the high-frequency information without frequency compression. The other two subjects showed a minimal benefit from the transposition, but only one of those subjects chose to use the Transonic as their main hearing aid. This does not suggest that the form of transposition provided by the Transonic is beneficial for improving speech perception scores. However, the study was based on a group of only five subjects, so further more extensive testing with this approach is required.

Posen et al. (1993) looked at the effectiveness of lowering the high frequencies of speech using a vocoder. The frequency lowering only occurred when the level of the high frequency energy was greater than that of the low frequency energy. The processing was compared to low-pass filtering with a cutoff frequency of 800 Hz and also to a vocoder system that frequency-lowered the entire speech spectrum. When testing normal-hearing subjects they found that the low-pass filtering elicited the poorest performance, frequency lowering where the whole spectrum was shifted down produced slightly better performance than the low-pass speech and the frequency-lowering approach which only shifted the frequency components down when the high frequencies were predominant elicited the highest scores of all. Another interesting finding was that performance of the subjects in all conditions improved over a period of ten sessions, showing that experience of the processing improved performance.

### *1.6.2 Speech pattern extraction*

A completely different approach to improving speech perception for severe-to-profoundly hearing-impaired listeners is that of speech pattern extraction (Fourcin et al., 1977). This aims to select those elements of speech that are most useful for speech perception, so as to simplify the speech signal. The elements provided are dependent on the individual's hearing abilities and are matched to the individual's frequency and intensity range. The first implementation of the speech pattern element approach ((Sinusoidal Voice) SiVo-1: Rosen et al., 1987) represented the voice fundamental frequency as an acoustic sinusoid. This version of the SiVo aid proved to be beneficial as an aid to lipreading because of the information it gave regarding intonation and voicing contrasts.

Additional features have since been added to the SiVo aid. These include speech amplitude envelope and voiceless frication information. Amplitude envelope information carries suprasegmental cues, and can enhance the perception of some manner and voicing contrasts. Voiceless excitation information provides an audible indicator of the presence of frication to assist with the perception of voiceless fricative and plosive consonants. One important aspect of this approach is that noise-resistant analysis methods can be utilized in the extraction of features, as has been implemented for the extraction of voicing and voice fundamental frequency (Walliker and Howard, 1990). Laboratory evaluations with English listeners (Faulkner et al., 1992), Chinese listeners (Wei, 1993) and European listeners (Faulkner et al., 1997) have shown that, at moderate signal-to-noise ratios, some listeners using the SiVo aid do slightly better than those with conventional aids. However in quiet many subjects with a severe or profound hearing loss still gain more information from a conventional aid than from the SiVo with this limited set of features. It is therefore necessary to increase the number of features on the SiVo to assess if a larger number of simplified features may lead to better performance than with conventional amplification.

Reed et al. (1991) derived a system of low-frequency artificial codes to represent different phonemes; they used two different upper cutoff frequencies of the stimuli; 300 and 500 Hz. The encoding approaches are listed below:

(1) Plosives and fricatives were represented by bandpass noise with a duration of 30 ms for plosives and either 100 or 200 ms for fricatives;
(2) Affricates were coded using two consecutive bands of noise with different centre frequencies and durations;
(3) Semi-vowels consisted of three-tone complexes of 100-ms duration;
(4) Nasals consisted of two-tone complexes of 50-ms duration;
(5) Voiceless consonants had a 100-ms gap between the vowel and consonant whereas the voiced consonants did not;
(6) 10 tone complexes carrying a frequency-compressed version of the natural

spectral envelope represented vowels and diphthongs.

Two different compression factors were used, 8.3 and 5, for the 300 and 500 Hz upper cutoff frequencies respectively. Fundamental frequencies of 30 and 50 Hz were used for the 300 and 500 Hz cutoff frequencies respectively.

Encoded stimuli were compared to low-pass filtered (300 and 500 Hz cutoff frequencies) stimuli in identification tests. They used normal hearing listeners. They found an improvement for the encoded stimuli over the low-pass filtered stimuli and they concluded that the scores obtained were typically superior to those for frequency-compressed speech with similar degrees of frequency shifting. The improvement was greatest for consonants.

It is possible that an approach which combines the speech pattern element approach and transposition could lead to better results than for either approach on its own. Possibly, simplifying the high-frequency spectral pattern and moving it down in frequency, combined with an unchanged low-frequency pattern could be useful. The most appropriate option would be dependent on the individual listeners' speech perceptual abilities. This approach is investigated in the current thesis.

## 1.7 Cochlear implants

Flynn et al. (1998) studied the speech perceptual performance of two groups of listeners, one with severe hearing impairment and the other with severe-to-profound hearing impairment. They compared their results to those for listeners tested by Skinner et al. (1994), who were cochlear implant users with the Nucleus-22 implant using the SPEAK processing strategy. On all speech tests, the subjects in the severe group performed at the same level as or better than the cochlear implant users. The severe-to-profound group performed at the same level as the implant users for CNC words (consonant-nucleus-consonant) and were significantly worse than the implant users for sentences in quiet and in noise. Flynn et al. concluded that many of the subjects in the severe-to-profound group would benefit from implantation, but stressed the crucial role of speech tests for selecting appropriate implant candidates, due to the wide range in performance of their listeners.

There is a general consensus that those hearing-impaired listeners with profound or total hearing loss who score less than 30% on open set speech perceptual tests with their hearing aids, will typically perform better with a cochlear implant than with conventional acoustic amplification. Furthermore, there is a trend towards implanting people with severe hearing losses who can gain some benefit from conventional amplification, scoring between 31-60% on open set aided sentence recognition tasks. There has to be some question as to whether these listeners could perform equally as well if the research were channelled into developing hearing aids most suited to these severe-to-profoundly impaired listeners. The majority of research with this group of listeners shows a large variability in the residual hearing abilities between the individual people. It would therefore seem appropriate to determine if other approaches to providing the speech information acoustically may be most appropriate for some people or whether the group as a whole is best suited by cochlear implants.

It is also known that if people use hearing aids and communicate orally prior to cochlear implantation (Osberger et al; 2002) that they will perform better post-cochlear implantation. There is therefore a need to provide people waiting for a cochlear implant with a hearing aid to stimulate the auditory cortex. Many current hearing aids are inadequate for severe-to-profoundly hearing impaired listeners and it is therefore essential to provide a hearing aid that cochlear implant candidates will utilize. The evidence also strongly indicates that the younger a child is implanted the more likely they are to acquire speech and language in the normal way (Sharma et al; 2002). The current trend in cochlear implants is therefore to implant children of 12 months or younger. This raises an interesting question for assessing the hearing abilities of these young children. It is practically very difficult to gain accurate results in conventional audiometry with infants so tests such as Oto-acoustic emissions and auditory brainstem responses are used, these are tests with limited benefits but they are accurate as the child is not required to respond and they depend on objective acoustic or electrophysiological measures. Oto-acoustic emissions measure the functionality of outer hair cells so can only give information upto a 60 dB HL cochlear loss and auditory brainstem responses measure responses of the auditory system at the level of the brainstem evoked by an auditory stimulus, this however only measures a limited frequency range but will give a good indication as to whether the child has an impairment or not. Although there are some limitations of the tests the risks are small compared to the risk of not implanting a child young enough.

## 1.8 Overview of the work in this thesis

The aim of the work described here is to determine if high-frequency frication cues to consonants, often inaudible for listeners with severe-to-profound hearing impairment, can be encoded to make them accessible to the hearing-impaired listener. One of the main areas of interest is whether place of articulation cues can be provided, to allow the listeners to understand this information without a heavy reliance upon lipreading. The work focuses on encoding voiceless fricatives, but it is hoped that the processing strategies used would also be appropriate for encoding other consonants.

The first experimental chapter assesses whether for normal-hearing listeners the voiceless excitation in fricatives can be simplified and represented by bands of noise. The second and third experimental chapters assess the abilities of a group of severe and profoundly hearing-impaired listeners to discriminate the spectral characteristics of bands of noise presented within their residual hearing range. The final experimental chapter describes an identification experiment using hearing-impaired listeners in which a range of processing strategies is employed for encoding voiceless excitation, to determine which approach would best encode the place cues.

# CHAPTER TWO

## 2 THE LABELLING OF BANDPASS FILTERED NOISE BANDS AS FRICATIVES BY NORMAL HEARING LISTENERS

### 2.1 Introduction

The main concern of this chapter was to determine if bands of noise varying in bandwidth centre frequency and relative level could represent the important perceptual cues carried by the frication section of natural voiceless fricatives.

In normal speech, there are many acoustic cues to signal the identity of a specific speech sound. It is essential to understand which cues are most important when considering an encoding scheme for hearing-impaired subjects. This is necessary because, although normal-hearing listeners can make use of a wide range of simultaneous acoustic cues and can compensate in difficult listening conditions by utilizing only a small subset of those cues, this is not always the case for hearing-impaired listeners, whose limited residual hearing abilities do not make all of the cues available, typically these would be the weaker higher frequency cues.

#### *2.1.1 Fricative perception*

Appendix 1 contains spectrograms of fricatives to assist with this review.

Harris (1958) conducted experiments to determine the most important cues used by normal-hearing listeners to identify fricatives. Vocalic and frication sections from the production of 16 different consonant-vowel (CV) syllables (C=/f, θ, ʃ, s/ & V=/i, e, o, u/) were spliced and recombined with each alternative. For example, for the /f/ and /i/ combination, there were four different /i/ vowels (from the /fi/, /θi/, /ʃi/ and /si/ productions) which were combined with the /f/ from the original /fi/ to give four stimuli. This was repeated for each fricative to give 16 stimuli arising from those uttered with an /i/ vowel. This was repeated for the stimuli with /e/, /o/, and /u/ endings to produce 64 stimuli in total. The stimuli were presented to 22 listeners in an identification task, to determine whether the noise-like portions or the vowel transitions were dominantly cues for the identification of the fricative sounds.

The results showed that the sibilant (/s/, /ʃ/) / non-sibilant (/f/, /θ/) distinction was determined by the intensity of the frication section of the syllable. For the sibilant consonants (/s/, /ʃ/), the centre frequency of the frication was used to determine the place of articulation whereas for the non-sibilant (/f/, /θ/) consonants, place was determined by the vowel transitions.

Heinz and Stevens (1961) synthesised a set of simplified fricatives with the purpose of determining if a series of simplified stimuli could generate appropriate responses from normally hearing listeners, and also to establish the ranges of spectral peak and valley combinations that would elicit appropriate responses for the individual fricatives.

The first stage of experimentation looked at the identification of the synthetic fricatives in isolation. The synthetic sounds were presented in isolation, with the resonant frequency varying between 2 and 8 kHz and with an anti-resonance at 1 octave below the resonance. Three bandwidths were used for each resonant frequency. For resonances at 6.5 or 8 kHz, additional low-frequency noise was added. Subjects had to label the stimuli as one of the following: /ʃ, ç, s, θ, f/.

The results showed that bandwidth had no significant effect so the results were averaged across bandwidth. As the resonant frequency increased, responses changed from /ʃ/ to /ç/ to /s/ to /f, θ/ (f and th could not be discriminated, so responses to these two were combined). The additional low-frequency noise had a marginal effect, increasing the identification of /f, θ/ but at the expense of the identification of /s/.

The second stage of the experiment looked at the identification of fricative-vowel syllables, using the vowel /a/. One set of stimuli did not contain formant transitions, and the other did. For those that did, only the first and second formants were present. The first formant transition started at 200 Hz, and the second formant transition started at 900, 1700 or 2400 Hz. There were 108 stimuli, which were presented twice to 8 listeners. The listeners had to identify each stimulus from a closed set of fricatives (/f, θ, s, ʃ/) during the first presentation and to rate the naturalness for the second. The results from this stage of experimentation showed that:

1) The presence of low-frequency noise gave a marginal improvement for the perception of /f and θ/;
2) If there were no transitions present, the result was generally worse than when there were transitions present;
3) Stimuli which were most strongly associated with a specific response were found to be the most natural;

4) /ʃ/ was always identified when the resonant frequency was around 2.5 kHz. /s/ was always identified when the resonant frequency was greater than 3 kHz, and /f, θ/ were identified when there was a very high resonant frequency, especially when low-frequency noise was present and of a low intensity;

5) /s/ and /ʃ/ sounded more natural if there were two spectral peaks, but /f/ and /θ/ only required one peak;

6) Formant transitions had a greater effect on the identification of /f/ and /θ/ than on /s/ and /ʃ/. This finding was in agreement with those of Harris (1958).

### *2.1.2 Audibility of cues*

The studies described so far have investigated the relevance of cues for normally hearing listeners. For hearing-impaired listeners other aspects have to be considered. Often the weak high-frequency cues are not audible for the impaired listener due to their limited frequency range. Also the audibility of those cues is low because of the listeners' raised thresholds. The stronger formants and formant transitions present in vowels are typically more audible than the weak high-frequency energy but it is possible that difficulties with vowel perception may arise due to impaired frequency selectivity.

Zeng and Turner (1990) tested four normally hearing subjects with natural and synthetic stimuli presented over a range of levels to determine the relative importance of the transition and frication sections of CV syllables in the identification of fricatives. At relatively low sensation levels, where they predicted that the frication was not audible, the subjects identified the fricatives with a reasonable degree of accuracy. This suggested that listeners could make use of the transition cue even at very low levels. Performance was also very good when the frication was audible but the transition cue was removed, indicating that the frication cue alone was sufficient to identify the fricatives. They concluded that the frication cue was the most important cue for fricative identification but the supplement of the more intense transition cue was of great benefit, especially at low levels when the frication was not audible.

### *2.1.3 Identification of fricatives by hearing-impaired listeners*

Zeng and Turner also investigated whether hearing-impaired listeners use the same acoustic cues for perception of the fricatives as normal-hearing listeners. They tested three hearing-impaired subjects, two with severe losses and one with a moderate loss. In a fricative-vowel (/s, ʃ, f, θ/) identification task, they found that when the formant transitions were made audible to their subjects, performance was not as good as for normal-hearing subjects given the same degree of audibility. Zeng and Turner's study also showed that if the high-frequency noise cues could be made audible to their subjects they performed similarly to a group of normal subjects at the same level of audibility. This finding suggests that some hearing-impaired listeners retain suitable residual abilities to discriminate the gross-spectral differences of the high-frequency cues in the obstruents. If it were possible to make these cues audible to hearing-impaired listeners this could be sufficient to help them perceive the obstruents. In contrast the formant transitions were found only to assist perception, and were not sufficient alone to give near-normal performance.

Lawrence and Byers (1969) tested a group of subjects with moderate sloping high-frequency hearing losses and found that they showed very good performance in labelling the fricatives /f, s, ʃ, θ/ in a CV context. They concluded that their subjects were able to make use of low-frequency acoustic cues. Such cues were potentially available in the noise spectrum below 2 kHz for some of the fricatives tested. They supported this conclusion with subjective reports of the subjects that they heard a buzz-like sound during the frication section of some of the fricatives. In addition these subjects, who had only moderate losses, may have made use of the formant transitions of the surrounding vowels to assist perception. This study is inconclusive in terms of knowing which cues the listeners with moderate hearing losses were using but it does suggest that the listeners were able to utilize the cues that were audible to them.

The studies described above suggest that if the aperiodic sounds present during fricatives can be made audible, then hearing-impaired listeners typically are able to utilise the information. This is most feasible for listeners with a moderate hearing loss with some residual hearing at high frequencies, for which frequency-response shaping to make the high-frequency cues audible should be achievable. However, for many people with severe-to-profound hearing impairment, it would be very difficult to make the high frequencies audible. Where there are dead regions at high frequencies, amplification may not provide audibility, and even where residual hearing is present in the high frequencies, it may not be possible to provide sufficient amplification.

## 2.2 Aim of current work

The work described in the present chapter is concerned with simplifying the high-frequency information of the obstruents and mapping it into the low-frequency residual hearing range of the severe-to-profoundly hearing-impaired listeners. Although this study only focuses on the voiceless fricatives, it is hoped that the results may be relevant to the other voiceless obstruents containing static aperiodic energy.

The main aim of this study was to assess if simple spectrally static band-pass noises, varying in centre frequency, bandwidth and intensity, but carrying no formant transition cues, could be labelled as fricatives by normally hearing subjects, based on acoustic cues similar to those relevant for the labelling of natural fricatives. If this turned out to be the case, it was planned to use the information to determine a range of simple stimuli for testing with hearing-impaired listeners.

## 2.3 Method

### *2.3.1 Subjects*

10 normally hearing subjects were tested in a sound-attenuating chamber. The subjects first underwent audiometric testing to ensure that their thresholds were within the normal range. All the subjects had some previous experience in speech perceptual testing and they were given an hour of practice prior to data collection.

### *2.3.2 Stimuli*

A series of synthesised stimuli was created. The stimuli consisted of a band-pass noise, with a 100-ms steady state portion and10-ms onset and offset ramps, preceded and followed by 200-ms bursts of a sinusoid with a fixed frequency of 100 Hz and 10-ms onset and offset ramps. The stimuli were intended to simulate the coding of an intervocalic consonant, the sine wave representing the vowel and the band of noise the fricative. The durations of the stimuli were based upon those observed for natural speech.

The simplicity of the stimuli was based on two considerations. Firstly, it was necessary that the relative importance of each cue could easily be assessed without interference from other cues. Secondly, the final aim of the project was to produce simplified stimuli for hearing-impaired listeners to improve speech perception, and it was essential to know if simplified cues could be used to represent aspects of individual phonemes.

The stimuli were created using the Klatt synthesizer (Klatt, 1980). They varied in: (1) centre frequency (CF), from 1 to 7 kHz in steps of 1 kHz; (2) -3 dB bandwidth, from 0.5 to 2 times the CF in steps of 0.5CF; (3) intensity of the noise band, from -15 to +6 dB relative to the sinusoid in 3-dB steps. The sinusoid had an overall level of 60 dB SPL. Not all the CF's could be employed for all bandwidths because, as CF increased, the bandwidth available was limited, due to frequency range restrictions (10 kHz) of the available version of the Klatt synthesizer. There were 176 stimuli in total.

The stimuli were created with a 20 kHz sampling frequency, stored on a Masscomp 5400 computer and played out through a 12 bit D/A converter and a Kemo VBF-8 anti-aliasing filter set at 10 kHz. Presentation was diotic through Beyer Dynamic DT48 headphones. Responses were collected via a Concept Keyboard that was labelled with the five response options (see below).

Appendix 2 shows an analysis of the rms level and the peak levels for the 0dB and –6dB stimulus groups. It shows how the rms level was equalised across the stimuli and that the peak level varied according to bandwidth and centre frequency due to the radiation characteristics of the Klatt synthesiser.

### *2.3.3 Task*

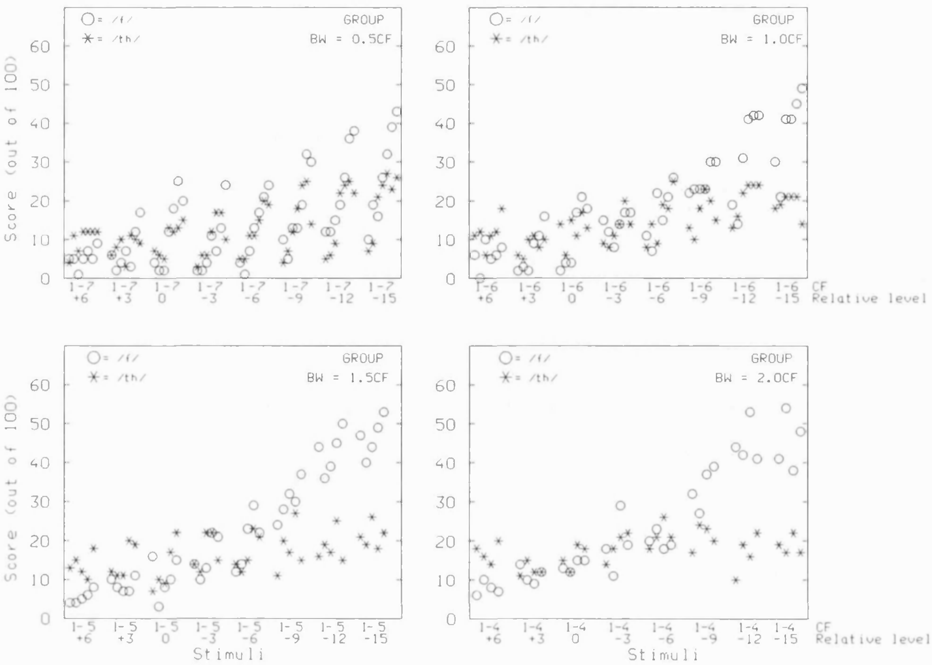
Subjects listened to 10 lists of stimuli, each consisting of the 176 different stimuli in a random order. Each list was presented in a different session. Subjects were requested to say which was the most appropriate label for the fricative-like part of each stimulus they heard, from the options /f/, /θ/, /s/, /ʃ/, /h/ and they were asked to respond even if they thought that the stimuli did not sound like any of the options. The stimuli did not sound like natural speech but the subjects were instructed to label the stimuli as speech sounds.

## 2.4 Results

The /h/ response label was seldom used, and when it was used it was not systematically associated with any of the stimuli. This is probably because the simple bands of noise used here completely lacked the distinctive formant structure typically occurring for /h/ in natural speech. It is possible that /h/ responses might be more readily elicited by lower frequency bands of noise or by multiple band noises.

The choice of label for the non-sibilants varied across individuals. Some subjects tended always to use the /f/ label, some used the /θ/ label, and others used a mixture of these two responses.

Figure 1 shows the percentage of times that responses of /f/ and /θ/ were given for each stimulus. Each panel represents a different bandwidth. Within the panels of figure 1, each column represents a different relative level, and within each column the CF increases

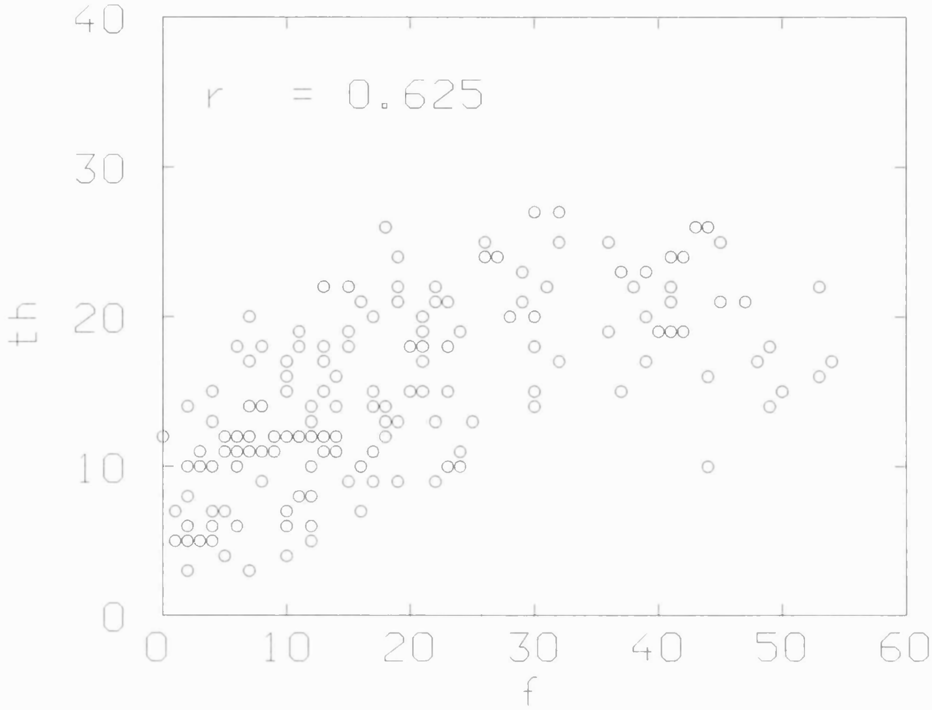



from the left to the right (numbers 1-7). Each column in figure 1 corresponds to a different relative level.

Figure 1. /f/, /θ/ scores out of one hundred for the group. Each panel refers to a different bandwidth. The x-axis shows the CF increasing from left to right, grouped by relative level.

The response rates for /f/ and /θ/ varied with CF and relative level in a similar way, except when the bandwidth was high and the relative noise level was low. For the stimuli with higher relative levels, the majority of subjects used the responses /f/ and /θ/ fairly indiscriminately. However, as the relative level was reduced, particularly for the broader bandwidth stimuli, subjects tended to prefer either one or the other response category. Most subjects preferred the /f/ label, one subject preferred the /θ/ label and two subjects still used both labels. A few of the subjects who preferred the /f/ label did show a CF-dependent response for their choice of labels, which could be connected with the fact that, in natural speech, /f/ tends to contain more high-frequency noise than /θ/. However, the overall group results did not show a dependence on CF of the relative response rates of /θ/ and /f/.

Figure 2 shows the relationship of the responses of /θ/ and /f/. Each point shows the cumulative score across all subjects for one stimulus. The /θ/ scores are plotted as a function of the /f/ scores. The Pearson's correlation coefficient is significantly different from zero ($r = 0.625$, $n = 176$, $p < 0.001$). Due to the high correlation of the /f/ and /θ/ scores,




the scores from /f/ and /θ/ were combined and will be referred to as /f/ from now on. All scores for the /h/ response (due to the small numbers) were divided equally between the other categories.

Figure 2. A function showing the relationship between /f/, and /θ/ responses for the group. The /f/ responses are plotted on the x-axis and the /θ/ responses on the y-axis.

Confusion matrices were created for each subject and the entire group, the responses for /h/ were divided between the other categories and the /f/ and /θ/ responses were combined. Figures 3, 4, 5 and 6 show the data for the entire group. Each response category is plotted with a unique symbol. The x-axis on each plot shows CF, the y-axis the percent score and each panel shows a different relative level of the noise. Figures 3, 4, 5 and 6 show results for bandwidths of 0.5, 1.0, 1.5 and 2.0CF, respectively. The scores for each stimulus were out of a total of 100 (10 subjects hearing each stimulus 10 times).

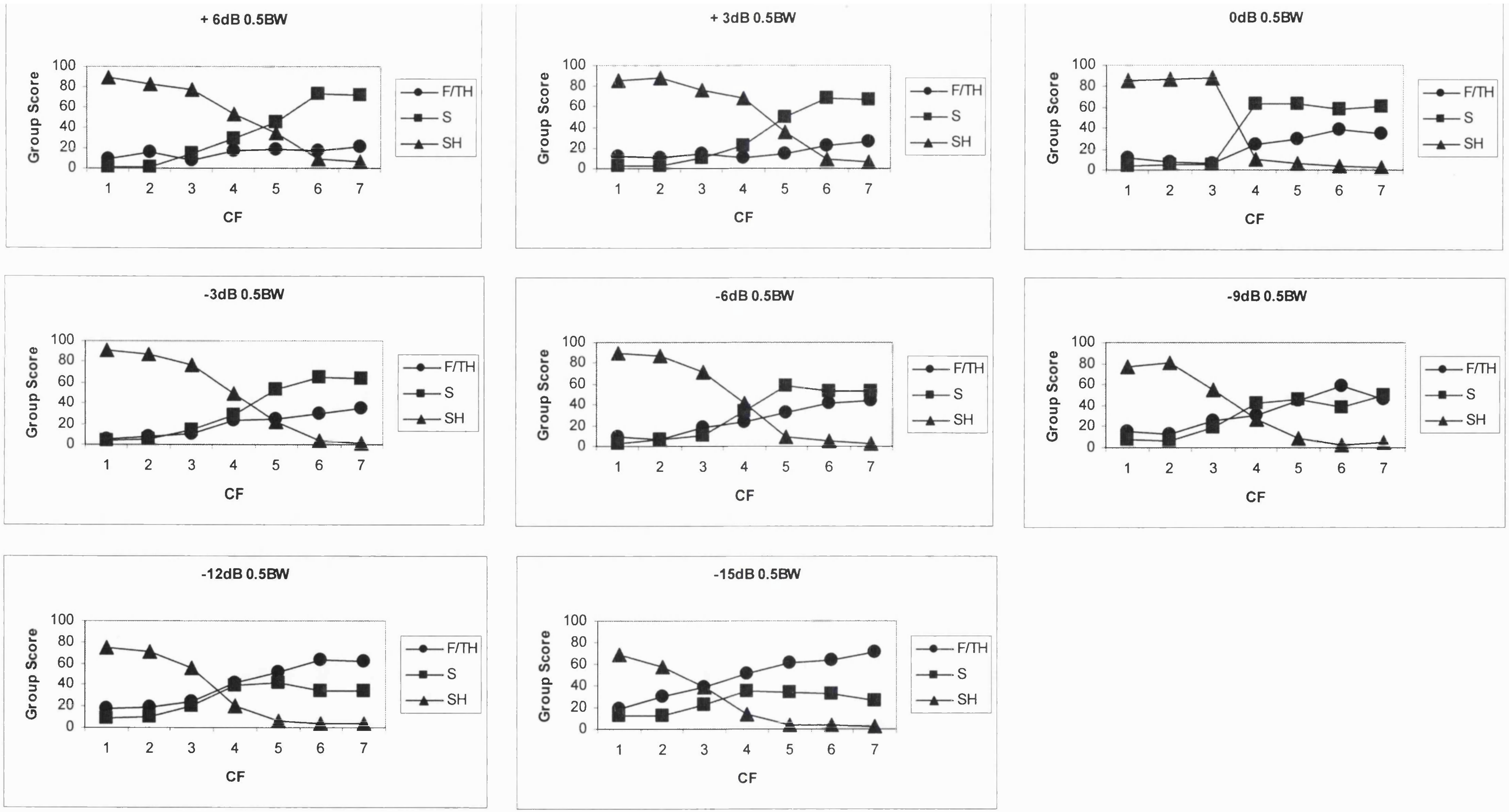



Figure 3. Percent response scores for the group when the bandwidth is 0.5CF. CF is on the x-axis and each individual plot refers to a different noise level relative to the sinusoid.

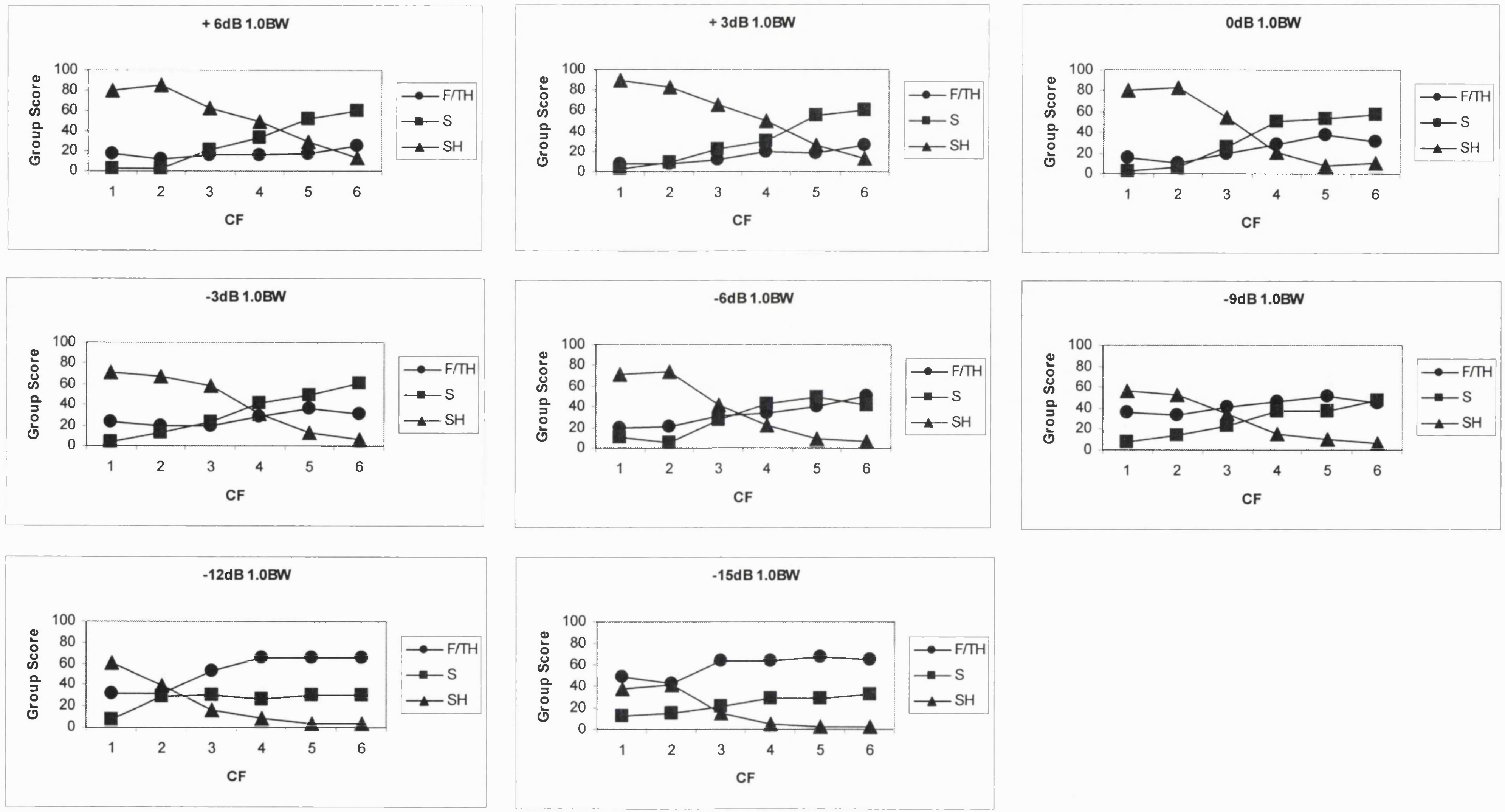



Figure 4. As for figure 3, except with a bandwidth of 1.0CF.

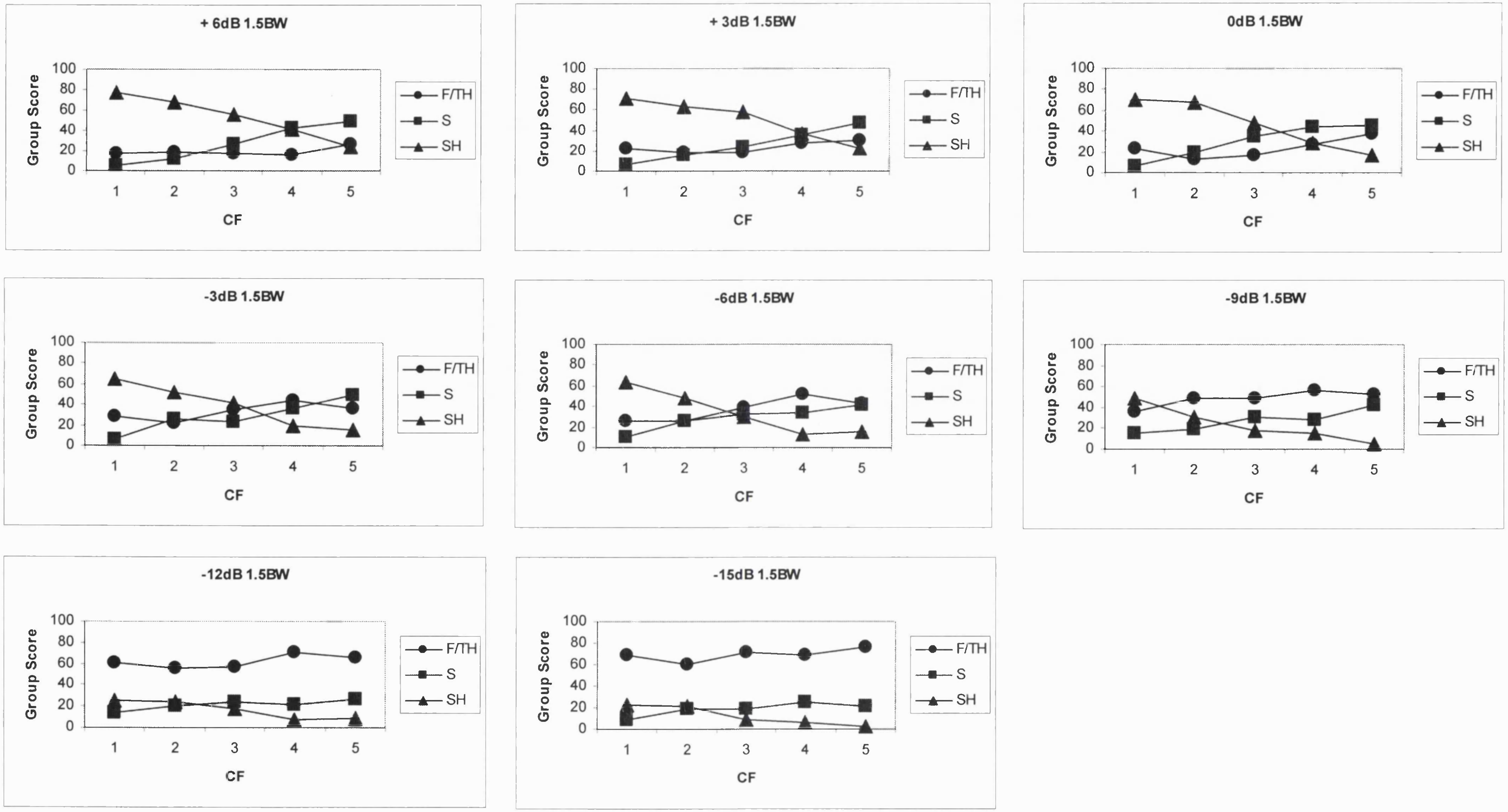



Figure 5. As for figure 3, except with a bandwidth of 1.5CF.

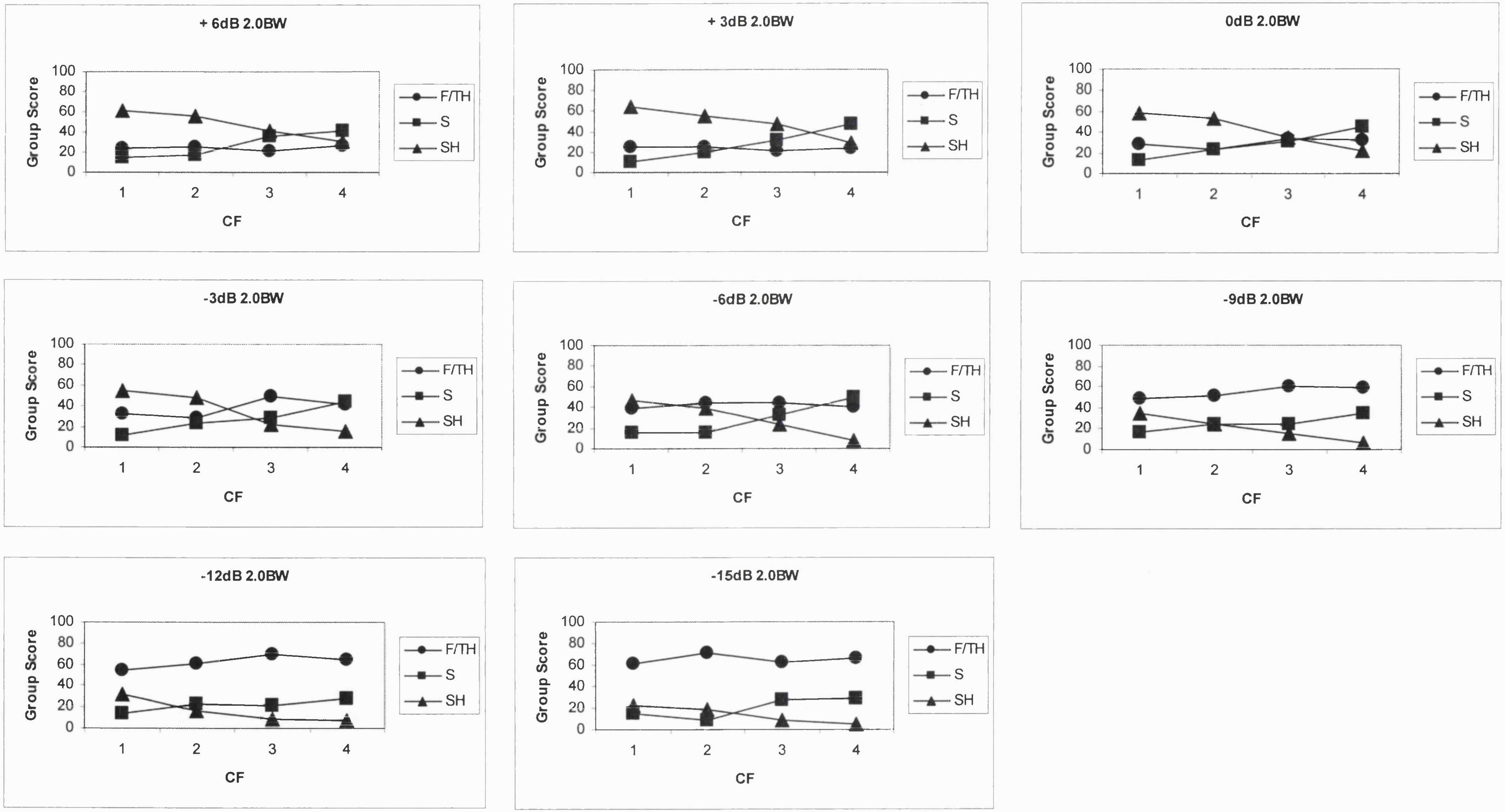



Figure 6. As for figure 3, except with a bandwidth of 2.0CF.

The sibilant response categories (/s/, /ʃ/) were typically used when the stimuli were high in relative level and had a narrow bandwidth. The sibilant responses showed an effect of CF. the /ʃ/ response was common for the low-frequency noise bands and the /s/ response was mainly given for the higher-frequency stimuli. There was an interaction of bandwidth with CF, because the /ʃ/ response was still a prevalent response when the bandwidth was increased to 2CF, again typically for the lower CF and higher-level stimuli, whereas the /s/ responses were dramatically reduced for the wider bandwidths, this of course could have been confounded by that fact that there were not such high CF's for the broader band stimuli due to the limitations of the Klatt synthesizer, however the high frequency components would still have been present in the higher CF stimuli of the broader bandwidth stimuli but they were not the CF. This suggests that the /ʃ/ response category may not be as sensitive to bandwidth as the /s/ category but to fully test this outcome higher CF's would need to be tested for the broader stimuli.

The results show that relative level was an important cue for differentiating the sibilants /s, ʃ/ and the non-sibilant /f/. The sibilant responses were given more frequently for the noise bands with high relative level and the bands of noise with low relative level were identified as non-sibilant more frequently.

Bandwidth was also an important cue for determining the difference between the sibilant/non-sibilant categories. There was a tendency for the narrower bands of noise to be labelled as the sibilants and the broader noises as the non-sibilants, although /ʃ/ responses were still frequently elicited for the broader noises at high relative levels.

Centre frequency also showed an effect, mainly for the sibilant categories, with the lower CFs being labelled as /ʃ/ and the higher CFs as /s/. CF was an effective cue for /f/ particularly when the bandwidth was narrow and the relative level of the noise was low; in this situation the noise bands with high CF's tended to be labelled as /f/.

Appendix 3 shows the mean standard deviations for the group as a function of response category and bandwidth, the standard deviations are small indicating a low inter-subject variability.

To summarise the results, /ʃ/ was a common response especially for stimuli which were low in frequency, high in relative level and narrow in bandwidth. The /s/ response was mainly given when the stimuli were of high frequency, high relative level and narrow bandwidth and the /f/ response was favoured for stimuli which were of high frequency, low relative level and had a broader bandwidth.

## 2.5 Discussion

The results of this experiment with normal-hearing subjects show that even very simple noises can carry some of the critical information required for labelling voiceless fricatives. It should be noted that because the subjects responded in a closed set it could not be determined if the responses to these stimuli would have been confused with responses to other possible speech sounds.

The fact that /h/ was seldom used as a label suggests that subjects were basing their responses upon similarities to natural speech. In natural speech, /h/ is typically more vowel-like, containing more spectral peaks than the other fricatives, and therefore a less appropriate label to give to the simple noise bands. In contrast, /f, θ, s, ʃ/ are spectrally less complex, with only one or two spectral peaks, and thus more similar to the stimuli used here.

The patterns for the /f, s, ʃ/ responses follow the general pattern of responses described by Harris (1958). The subjects tested here could label the sounds as one or the other sibilant based on the static noise characteristics but they did not differentially label the non-sibilants with these static cues. When the noise was low in relative level, contained high-frequency components and had a broad bandwidth, this was sufficient to give an impression of non-sibilance. There was not, however, sufficient information to signal non-sibilant place of articulation. Hence some subjects preferred the /f/ and some the /θ/ label, supporting the idea that formant transitions play a big role in cueing non-sibilant place information. The choice of the sibilant/non-sibilant categories was based on the relative level and bandwidth of the noises. The narrow bandwidth sounds with high relative level were generally given sibilant labels and the broader noise bands with low relative level were labelled as non-sibilant. /f/ was typically given as a response for high-frequency tokens with the latter characteristics. Mainly the CF of the noise band determined sibilant place of articulation. Low frequencies were labelled as /ʃ/ and high frequencies as /s/.

The pattern of results in many ways reflects predictions that would be made based upon spectrographic analysis of natural fricative utterances. Strevens (1960) in a summary of the spectrographic analysis of voiceless fricatives summarized the productions of /f/, /s/, /ʃ/ as:

1) /f/ contains low level energy at around 1500 – 1700 Hz and extending higher than 7500 Hz. Peaks often occur around 1900, 4000 and 5000 Hz.

2) /s/ has the lowest frequencies above 3500 Hz. It contains high-level energy extending high in frequency to above 8000 Hz.

3) /ʃ/ The lowest frequencies vary between 1600 and 2500 Hz and can extend up to 7000 Hz, however the highest level regions tend to be in the lower frequencies. It also contains high-level energy.

The sibilants typically contain high-level frication sections compared to the low-level frication present during the production of the non-sibilants. The level of the noises in the labelling experiment had a large effect on whether the tokens were labelled as sibilant or non-sibilant. This was especially the case for /f/ versus /s/ as these were both labels given for noises with high centre frequencies, but differentiated based on the relative level. Centre frequency played a role perceptually for the differential labelling of /s/ as opposed to /ʃ/, which fits in with the results of spectrographic analysis which indicate that /ʃ/ has more low frequency energy than /s/. Strevens (1960) spectrographic analysis indicated that /f/ had the widest bandwidth for the spread of energy (1700 - > 7500 Hz), /ʃ/ was also fairly wide (2500 – 7000 Hz) but /s/ was relatively narrow (3500 - >8000 Hz). The results also follow this pattern of production with /s/ being most regularly given as a label when the bandwidth was narrow and /f/ and /ʃ/ being most commonly given as responses for wider bandwidth sounds, /f/ being more commonly given in the widest bandwidth category.

### *2.5.1 Implications for speech coding*

The results of this experiment are encouraging because they suggest that simple noise bands can be used to represent fricatives. This has implications for the encoding of voiceless fricatives for severely and profoundly hearing-impaired people. Many of these people cannot make use of formant transition cues and it can be very difficult to make the noise-like section of fricatives audible to them, due to the severity of their hearing loss in the high frequencies. One approach to overcoming the problem would be to map the high-frequency acoustic energy into the low-frequency hearing range, where hearing-impaired listeners usually retain some useful discrimination abilities. The potential problems with this approach are two-fold. Firstly, if the spectral peaks in the stimuli are frequency compressed, maintaining their frequency ratios, it is possible that the peaks would be too closely spaced for an impaired auditory system to resolve. Listeners with normal hearing or a moderate loss at low frequencies might benefit from this approach, but it is unlikely that listeners with severe losses at low frequencies would have sufficient frequency selectivity to resolve the compressed peaks. Secondly, it is possible that, once the stimuli are transposed down in frequency to within the residual frequency range of hearing-impaired listeners, the cues that normal hearing listeners use to label noise bands may be altered so much that they are not useable. This caution particularly applies to the effect of centre frequency upon the discrimination between /s/ and /ʃ/. Once lowered in frequency, it is possible that all of the sibilant representations would be identified as one of the sibilants. However, it is predicted that as long as the representations can be discriminated, they should be able to represent the important perceptual cues. It may be necessary to give auditory training to allow people to label the encoded stimuli.

Rosen et al. (1999) showed that normal hearing listeners could dramatically improve their speech perception performance when listening to speech containing spectral cues shifted up in frequency after only 3 hours of training using connected discourse tracking (De Fillippo and Scott, 1978). The training was split up so that after the initial assessment, 20 minutes of training was given in each of nine sessions. The frequency shifts were equivalent to a 6.46 mm displacement along the basilar membrane giving a 1.3 - 2.9 octave shift in frequency, depending on the frequency region. The simulations were based upon a four channel cochlear implant processing. The original speech signal was separated into 4 bands and the envelope extracted. The envelope in each channel was then multiplied with white noise and the outputs filtered prior to being added together. In the spectrally shifted conditions the edge frequencies of the output filters were increased slightly to simulate the 6.46 mm shift. Initially the mean BKB (Bamford-Kowel-Bench) score for shifted sentences was 1% as opposed to 64% for unshifted sentences. The scores for the shifted condition increased to 30% in the final session. Their subjects were not exposed to the speech on a daily basis, so these results could possibly be improved upon. This shows that people can learn to use new speech representations when the sounds are placed in different spectral locations.

The stimuli used in this experiment were not derived from natural speech tokens. The experiment was conducted to determine if simple noise bands could carry relevant information for labelling fricatives and to determine which acoustic cues were most important. For encoding fricatives for hearing-impaired subjects an approach that tracked the dynamic changes in speech would be necessary. A systematic approach to the encoding of fricatives by relating the encoded versions to the bandwidth, intensity and main peak frequency of the original frication would be most appropriate.

# CHAPTER THREE

## 3 THE DISCRIMINATION OF SPECTRALLY ADJACENT BANDS OF NOISE BY HEARING-IMPAIRED LISTENERS

### 3.1 Introduction

Due to their limited auditory frequency range, severe-to-profoundly hearing-impaired listeners can seldom use high-frequency cues for speech perception. In particular, the voiceless obstruents (fricatives, plosives and affricates) have high-frequency aperiodic components that contain crucial cues to place and manner of articulation. However, the majority of these people have potentially useful auditory abilities that are not exploited by conventional amplification. By simplification and transposition it may be possible to encode key features of these sounds within a limited residual hearing range. Faulkner *et al.* (1990) showed evidence of frequency selectivity at 125 and 250 Hz with estimated filter bandwidths 2 to 3 times greater than normal. This considerable broadening of filter bandwidth causes difficulties in resolving the relatively complex spectral structure of speech but may not prevent the spectral discrimination of simple speech-related stimuli. Rosen *et al.* (1990) determined that many of these listeners could discriminate periodic from aperiodic stimuli. These abilities indicate a potential for the speech pattern element approach (Rosen and Fourcin, 1983) in encoding the spectra of voiceless obstruents. The first stage of developing such an encoding method was to investigate the basic perceptual abilities of these listeners for discriminating noise spectra.

There has been very little research examining the discrimination of the spectrum of bands of noise in normal hearing or hearing-impaired listeners. Farrar et al. (1987) looked at the discriminability of speech-like noise stimuli based upon the spectra of noise-excited components in voiceless plosives and fricatives. They tested normal-hearing listeners. The stimuli were all broadband signals varying in their spectral shape according to the speech token being modelled. Within a run, two tokens were discriminated and the difficulty of the task was adjusted by adapting the level of a noise with the same long-term spectrum as the speech from which the tokens were extracted. They found that, as duration of the stimuli increased, the discriminability of the target noises improved.

Pickett et al. (1965) measured just-noticeable differences (jnds) for the cut-off frequency of low-pass noise with a variety of cut-off frequencies (250, 500, 1000, 1500 and 2000 Hz) in normal and severely cochlear-damaged subjects. At 250 Hz they found that the jnd for normally hearing subjects was 12%, which represents a 30 Hz difference. At the higher frequencies, their jnds ranged from 3 to 5%. The Pickett et al. study also showed that the difference between the normal hearing and hearing-impaired groups of subjects was not large at 250 Hz. The jnd for the hearing-impaired subjects was 15% at 250 Hz, which corresponds to an absolute frequency difference of 37.5 Hz. Performance worsened at higher cut-off frequencies for the hearing-impaired group probably because of a greater degree of hearing-impairment in the high frequencies and possibly the presence of dead regions (Borg et al. 1995).

Versfeld (1997) conducted an experiment to determine how well normally hearing listeners could detect changes in the spectral shape of noise bands centred at 1 kHz. Bandwidths ranged from 0.5 to 24 semitones (ST) (29 to 1392 Hz). Subjects were required to discriminate noises with a positive sloping spectral envelope from those with a negative sloping spectral envelope. He found that performance was best for bandwidths of 3-6 ST (174 to 348 Hz). Thresholds increased somewhat for wider bandwidths and very dramatically for bandwidths below 1 ST (58 Hz). Versfeld suggested that, for bandwidths below 3 ST, subjects were listening through one auditory filter and therefore were using temporal cues and reported using a change in pitch to discriminate the stimuli. For bandwidths greater than 3 ST, he suggested that subjects were making across filter (channel) comparisons, which may be why listeners reported listening to changes in timbre. For the larger bandwidth noises the edge bands were most appropriate.

The results using stimuli with frequencies below 4 kHz are of interest for the current work because the hearing-impaired listeners used in these studies typically have no hearing at high frequencies. In summary the studies described here suggest that:

1) Jnds using noise stimuli are typically much worse than those observed for pure-tone stimuli in the literature.
2) For noise bandwidths that are less than one auditory filter bandwidth, the ones with the relatively wider bandwidths typically result in the poorest performance.
3) For stimuli with low centre frequencies (4 kHz and below), listeners were more reliant compared to the higher frequencies, upon temporal information to discriminate noise frequency.
4) When the bandwidth of the noises is wider than the auditory filter bandwidth then listeners use across channel cues for discrimination.
5) For low-frequency stimuli, there is little difference between normal-hearing and hearing-impaired listeners.

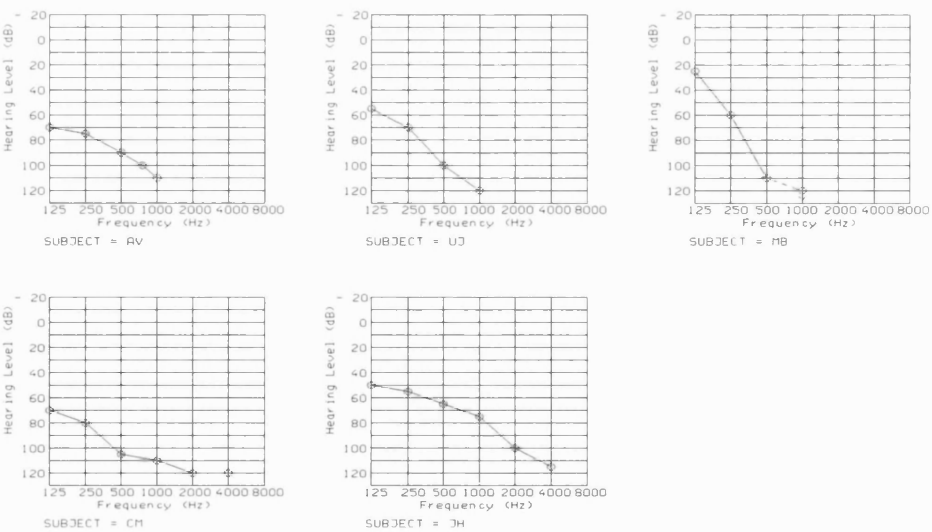
## 3.2 Rationale

The aim of the work described in this chapter was to determine the minimum duration at which hearing-impaired listeners could discriminate spectrally adjacent noises of the same bandwidth. The purpose of this was to assess if there was potential for a speech pattern recoding strategy, where the spectra of noise-excited consonants were represented by band-pass noises matched to the low-frequency hearing abilities of the listener. For such a strategy to be feasible, spectral discrimination would have to be possible for durations comparable to those of voiceless fricatives and plosives.

## 3.3 Method

### *3.3.1 Subjects*

Ten severe-to-profoundly hearing-impaired listeners were tested. The subjects were aged between 35 and 75 and all had post-lingual sensorineural loss. All subjects had previous experience in psycho-acoustic and speech perceptual experiments. Audiograms for the subjects are shown in Figure 1. The circles show the measured hearing thresholds; the circles with downward-pointing arrows indicate that thresholds could not be measured at the limit of the audiometer; the limit for that frequency is the value shown. Audiologists at the Royal National Throat Nose and Ear Hospital, London measured the hearing losses.



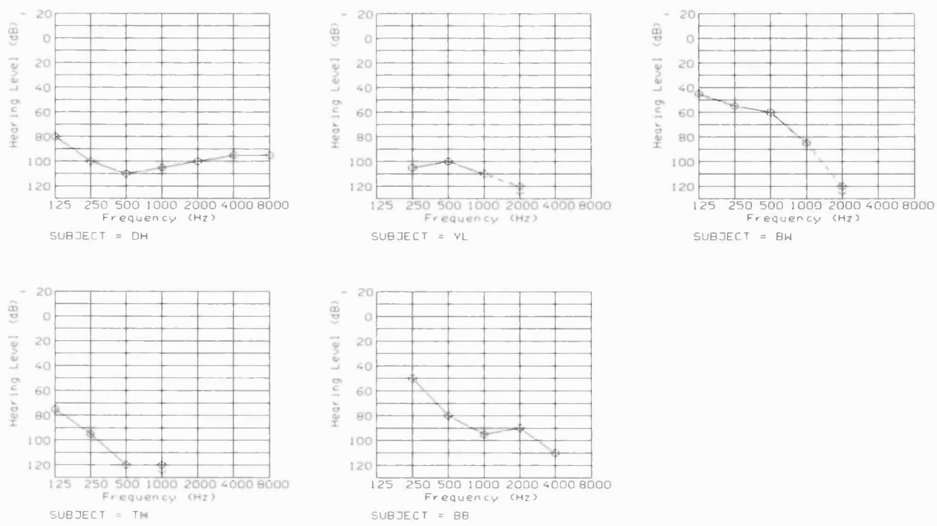



Figure 1. Audiograms for the ten subjects participating in the experiment

### 3.3.2 *Description of audiometric losses*

Three subjects had severe losses, with three-frequency average losses (at 500, 1000, and 2000 Hz) between 80 and 95 dB HL. The other seven subjects had profound losses, the three-frequency average being between 95 and 120 dB HL. The three-frequency average losses are shown in table 1. If a value was not measurable, 120 dB was used.

| Subject | 3FA 500, 1000, 2000 Hz | Subject | 3FA 500, 1000, 2000 Hz |
|---|---|---|---|
| TH | 120.0 | UJ | 106.7 |
| YL | 110.0 | AV | 106.7 |
| CM | 111.7 | DH | 105.0 |
| MB | 116.7 | BB | 88.3 |
| BW | 91.7 | JH | 80 |

Table 1. Three-frequency average hearing levels for the ten subjects.

### 3.3.3 *Task*

Subjects were required to discriminate two bands of noise of the same bandwidth, where the upper cut-off frequency of the lower-frequency noise band was equal to the lower cut-off frequency of the higher-frequency noise band; this common cutoff frequency will be called the "common edge frequency". Within a run, the common edge frequency and the bandwidth were fixed and the duration of both noises was adaptively changed. Two duration criterion values were used to interpret the data. These criterion values were based upon the shortest observed plosive and fricative durations in running speech (Howell and Rosen, 1983), 30 and 80 ms respectively. The procedure was repeated for a wide range of spectrally adjacent noise stimuli. A schematic illustration of the task and stimuli is shown in figure 2.

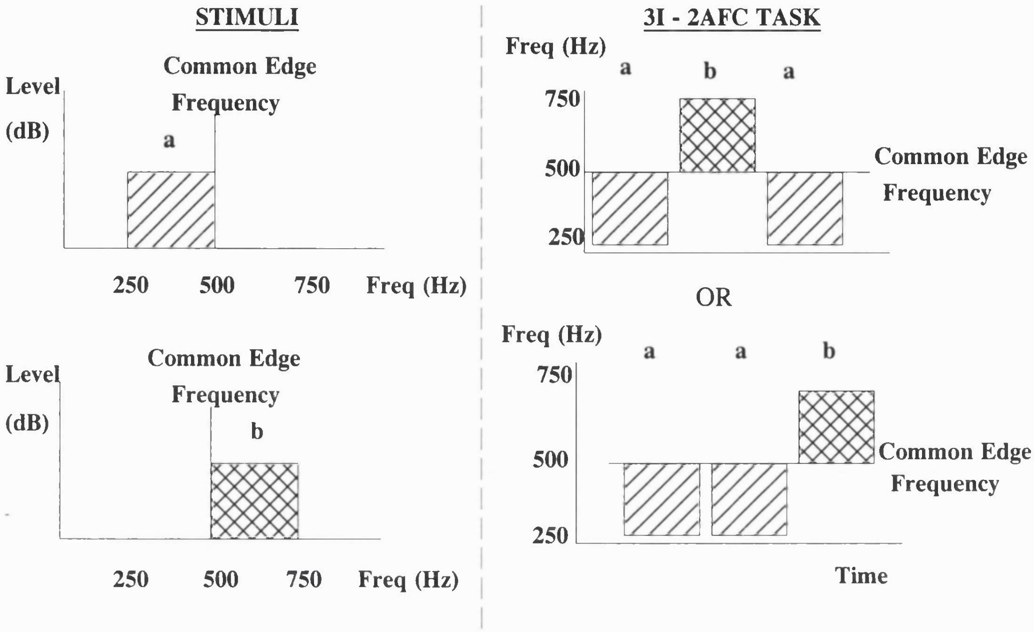



Figure 2. A schematic diagram of the stimuli and task.

The common edge frequency ranged from 125 Hz to 2000 Hz and bandwidths of 125, 250, 500 and 1000 Hz were used. Testing always began with the lowest common edge frequency available for a specific bandwidth, (e.g. 125 Hz for a bandwidth of 125 Hz). The common edge frequency was increased in integer multiples of the bandwidth for consecutive runs (e.g. 125, 250, 375 Hz were the three lowest common edge frequencies for the 125-Hz bandwidth), until subjects could no longer achieve a duration threshold below 80 ms. The full range of bandwidths could not fully be explored with all subjects because of limitations of the subjects' time and availability and also limits to their auditory ability.

Testing was carried out using a three-interval two-alternative forced choice paradigm, in which either the second or the third stimulus was higher in frequency than the first. The common edge frequency and bandwidth were fixed and the stimulus duration was varied. Subjects were asked to decide whether the second or third sound was the odd one out. The duration was adaptively altered according to a three-up, one-down staircase procedure (Levitt, 1971) converging at the duration giving 79% correct discrimination. The duration was changed by a factor of 1.35 for the first three reversals, and by a factor of 1.2 for a further ten reversals. The geometric mean duration at the last ten reversals was used as the threshold estimate. Lights in the response keys provided trial-by-trial feedback.

Subjects were tested in a sound-attenuating chamber and received a minimum of two hours practice before data were collected.

At least two thresholds were used to calculate the individual subject means for each condition. If the difference between those two means was more than 15 ms then two further means were measured and all four results were used. If the second set of two thresholds did not fall within the range of the first two, then further measurements were made until four consecutive thresholds were obtained that fell within a 20-ms range.

### *3.3.4 Stimuli*

The noise stimuli were derived from a white-noise source, and band-pass filtered using two Kemo VBF/8 filters, giving an overall slope of 90 dB/octave. Each stimulus was gated on and off with a half cycle of a raised-cosine function lasting 5 ms. This limited the minimum stimulus duration to 10 ms. The maximum duration was 600 ms. A Masscomp 5400 computer was used to control stimulus presentation through 12-bit D/A converters. A Yamaha amplifier, Hatfield manual attenuators, and Charybdis programmable attenuators were used to control the level of the stimuli.

Stimuli were presented monaurally through Beyer Dynamic DT48 headphones; electronic equalization was applied digitally via a Barr and Stroud EF8 digital filter to give a flat headphone frequency response down to about 70 Hz. A microphone mounted on the headphone grille and an FFT analyser (Ono-Sokki CF-910) were used to monitor the signal levels and spectra.

Stimuli were presented at a comfortable listening level chosen by the subject. In order to eliminate frequency-related loudness cues, the different noise stimuli were balanced in loudness. It was essential to remove loudness cues to ensure that subjects were not using loudness differences to perform this psychoacoustic task. However for a final selection of stimuli to be used for speech perception, loudness cues provide useful information for discriminating between different speech sounds.

### *3.3.5 Loudness balancing*

The balancing procedure required the subject to rotate a potentiometer wheel, whilst hearing the two noises in alternation. The potentiometer adjusted the level of one noise while the level of the other noise was fixed. The subject was instructed to adjust the wheel and find the point at which the two sounds appeared to be equally loud. Each subject was advised to pass through the equal loudness point a few times to ensure accurate balancing.

This procedure was carried out three times prior to each set of threshold measurements for a given noise bandwidth and common edge frequency. If the three balance points were within a 3 dB range, the mean was calculated and used as the appropriate level. If the range was greater than 3 dB, another three measurements were taken and the mean of all six was used to determine the matching level. The subjects listened to the balanced stimuli to check they were at a comfortable listening level after one of the stimuli had been adjusted. If necessary the output level was adjusted. In order to eliminate any residual loudness cues due to balancing errors, levels were randomly varied over a 3 dB range from stimulus to stimulus within each trial.

## 3.4 Results

The results are shown in figure 3. Each panel shows results for one subject, with the duration threshold plotted on the y-axis and the common edge frequency on the x-axis. Note the criterion lines indicating the 80 and 30 ms values. It should also be noted that any points plotted at 140 ms, indicate thresholds of 140 ms and over, indicating that subjects did not perform well in that particular condition.

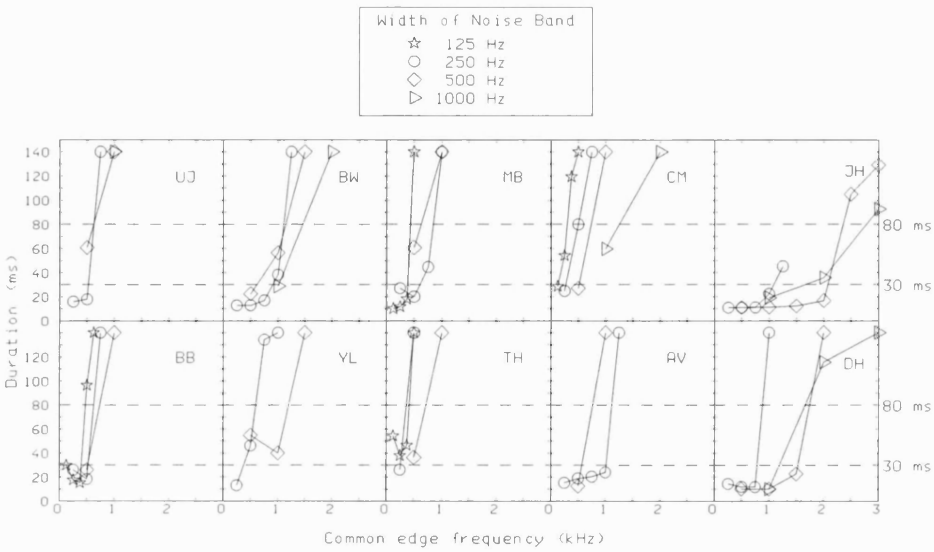


Figure 3. The duration thresholds of all ten subjects for a range of stimulus bandwidths.

Each set of data within the individual subject panels relates to a different noise bandwidth. There are different symbols for each bandwidth: stars for 125 Hz; circles for 250 Hz; diamonds for 500 Hz; and triangles for 1000 Hz. Not all of the subjects were tested in all of the different bandwidth conditions, but they were all tested at least for bandwidths of 250 and 500 Hz. If time permitted, the subjects with wider frequency ranges were tested with 1000-Hz wide stimuli and, if possible, 125-Hz wide stimuli. For the subjects with more limited frequency ranges, testing was carried out for the 125-Hz bandwidth if there was time. Only one subject (CM) was tested for all four bandwidths. All subjects obtained duration thresholds lower than 30 ms in at least one bandwidth condition. Typically, performance got worse as the common edge frequency increased and in general there was a sharp cutoff beyond which subjects could not perform the task.

For discrimination to be relevant to speech contrasts, an ability to discriminate centre frequency for durations less than 80 ms is required. As explained above, the two duration criterion values chosen were 80 and 30 ms.

Table 2 summarises the results from figure 3. For each subject, it shows the common edge frequency up to which the threshold was below 80 or below 30 ms. The inferred number of discriminable noise bands, extrapolated from this two-noise-band discrimination task, is also shown for the criterion durations of 30 and 80 ms, for each bandwidth and each subject.

To predict the number of discriminable noise bands, it was assumed that the percept, which allowed discrimination in this task, was monotonically related to frequency. This was supported by the listeners reporting that higher centre frequency noises sounded "higher". Thus if noise x was discriminated from a spectrally adjacent noise y and y was discriminated from noise z, it was assumed that x would be discriminated from z and thus that all three noise bands were discriminable.

| | Bandwidth | Highest Common edge frequency (Hz) -30 ms | Inferred no. of discriminable noise bands | Highest Common edge frequency (Hz) - 80 ms | Inferred no. of discriminable noise bands |
|---|---|---|---|---|---|
| BB | 125 | 375 | 4 | 375 | 4 |
| | 250 | 500 | 3 | 500 | 3 |
| | 500 | 500 | 2 | 500 | 2 |
| YL | 250 | 250 | 2 | 500 | 3 |
| | 500 | NA | 0 | 1000 | 3 |
| CM | 125 | 125 | 2 | 250 | 3 |
| | 250 | 250 | 2 | 500 | 3 |
| | 500 | 500 | 2 | 500 | 2 |
| | 1000 | NA | 0 | 1000 | 2 |
| JH | 250 | 1000 | 5 | 1250 | 6 |
| | 500 | 2000 | 5 | 2000 | 5 |
| | 1000 | 1000 | 2 | 2000 | 3 |
| TH | 125 | NA | 0 | 375 | 4 |
| | 250 | 250 | 2 | 250 | 2 |
| | 500 | NA | 0 | 500 | 2 |
| AV | 250 | 1000 | 5 | 1000 | 5 |
| | 500 | 500 | 2 | 500 | 2 |
| DH | 250 | 750 | 4 | 750 | 4 |
| | 500 | 1500 | 4 | 1500 | 4 |
| | 1000 | 1000 | 2 | 1000 | 2 |
| UJ | 250 | 500 | 3 | 500 | 3 |
| | 500 | NA | 0 | 500 | 2 |
| BW | 250 | 750 | 4 | 1000 | 5 |
| | 500 | 500 | 2 | 1000 | 3 |
| | 1000 | 1000 | 2 | 1000 | 2 |
| MB | 125 | 375 | 4 | 375 | 4 |
| | 250 | 500 | 3 | 750 | 4 |
| | 500 | NA | 0 | 500 | 2 |

Table 2. The highest common edge frequency at which two bands were discriminable for criterion durations of 30 and 80 ms, and the inferred number of discriminable bands of noise for each bandwidth and subject. For some subjects, with some noise bandwidths, discrimination was possible at durations below 80 ms but not below 30 ms. In these cases, NA (not applicable) is written in the 30 ms column, and the expected number of discriminable bands is zero.

It is important to note that it is the common edge frequency that is specified in the table. Therefore, if the bandwidth of the noises is 250 Hz and the highest common edge frequency is 750 Hz, this indicates that a 500 - 750-Hz noise can be discriminated from a 750 - 1000-Hz noise below the specified duration criterion. This example would suggest that four successive bands of 250-Hz wide noise could be discriminated by the subject.

All subjects were capable of discriminating noises in at least one condition. The overall pattern of results showed that performance deteriorated with increasing common edge frequency and also with increasing bandwidth. For the 250-Hz bandwidth, every subject could discriminate at least two noise bands at durations of 30 ms or less. For those subjects who were tested with the 125-Hz wide noise bands, performance was similar to or better than that for the 250-Hz bandwidth (with the exception of subject TH). This indicates that subjects can be expected to discriminate amongst twice as many adjacent bands of 125-Hz wide noise than of 250-Hz wide noise. All the subjects were tested with the 500-Hz wide stimuli. Discrimination for at least one common edge frequency below the 80 ms criterion was possible for all subjects, but some subjects (MB, UJ, TH and YL) could not discriminate any of the noises at the 30 ms criterion. Four of the subjects (BW, JH, DH and CM) were tested with the 1000-Hz-wide noises. Three of these (BW, JH and DH) could discriminate for at least one common edge frequency below the 30 ms criterion and all four subjects were able to discriminate for at least one common edge frequency below 80 ms. Subject JH was able to discriminate 1000-Hz-wide noise bands for two common edge frequencies with a duration threshold below 80 ms.

Figure 4 shows a scatter plot of absolute thresholds versus duration thresholds for a specific frequency/common edge frequency. Only values where there was a measured threshold on the pure tone audiogram are shown. Duration threshold is shown on the logarithmic y-axis and hearing threshold is shown on the x-axis. A logarithmic scale for duration was used as this gave a more linear relationship between absolute threshold and duration threshold.

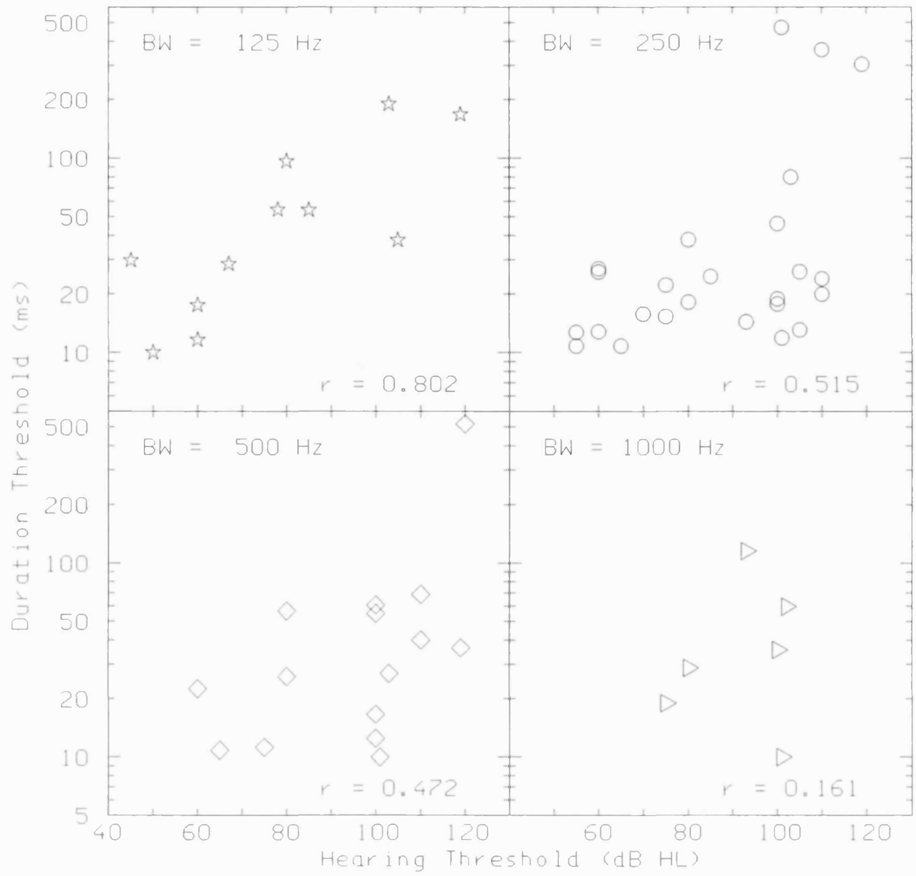



Figure 4. Plot of absolute threshold versus duration threshold (log scale). Each point is for a specific frequency/common edge frequency. Each panel shows results for one bandwidth.

For the 125-Hz bandwidth, duration thresholds show a reasonably linear relationship to absolute pure-tone 3FA thresholds. The data for 250-Hz bandwidth are also approximately linear with absolute threshold, but there is more scatter than for the 125-Hz bandwidth at absolute thresholds above 100 dB HL. The results for the 500-Hz and 1000-Hz bandwidths show increasing degrees of scatter.

Pearson correlation coefficients were compared to zero using t-tests; a 1% significance level was used to avoid false positives because of the number of t-tests being conducted. The results indicated that the Pearson correlation coefficients for 125- and 250-Hz bandwidths were both significantly greater than zero ($t (9) = 0.403$, $p>0.005$ and $t (22) = 2.819$, $p>0.005$, respectively) and those for the 500- and 1000-Hz bandwidths were not. The slope for the 250 Hz condition may have been affected slightly due to the outliers but it is unlikely to have affected the highly significant result. Tests were also carried out to see if there were any significant correlations between the three-frequency average hearing loss and number of discriminable bands, for the 30 and 80 ms criterion values. None of these correlations were significant.

The fact that there was a significant correlation between duration thresholds and hearing loss at the common edge frequency for the narrower bandwidths but not the wider bandwidths indicates that the cue used was nearer to the common edge frequency in the narrow band noises than the wider noises. One possible explanation is that subjects were using cues to the centre frequency of each noise band, which for the narrower bands is nearer to the edge frequency than for the wider noises and also for narrow bands the centre frequency is more strongly represented.

### 3.5 Discussion

The results showed that all subjects were to some degree capable of performing the task for durations of 80 ms or less and the majority of subjects could perform the task for some of the noise bands for durations below 30 ms.

All subjects participated in the 250- and 500-Hz bandwidth conditions. For the 250-Hz bandwidth, all the subjects were able to perform the task for at least up to one common edge frequency, at the 30 ms criterion value. For the 500-Hz bandwidth, all subjects could perform the task for at least one common edge frequency at durations below 80 ms, but below 30 ms four of the subjects could not perform the task at all. The greater difficulty for the 500 Hz bandwidth could have occurred because the audible frequency range for some subjects was narrow, with the result that the upper edge was inaudible. Also, if the subjects had dead regions at the high frequencies, this could have distorted the perception of the noise because neurones on the edge of the dead region would have been encoding all the information. Finally, as the bandwidth of a noise increases it is more difficult to determine the centre frequency of the noise and so it may be difficult to perform the task with only a 30 ms sample of the noise.

Four subjects were tested with the 125-Hz wide noise bands and all subjects were capable of discriminating up to two or three common edge frequencies for the 80-ms criterion. Three of the subjects could discriminate at least for one common edge frequency to the 30-ms criterion value and the other subject (TH) could not perform the task at this criterion. It is possible that, due to the high inherent fluctuations of the narrower noises, subjects like TH, who have a narrow dynamic range, could have set the comfortable listening level slightly too low to avoid any discomfort, and thus at very short durations audibility affected the overall performance.

The common edge frequency up to which discrimination was possible within the criterion limits decreased as the noise bandwidth decreased, i.e. the narrower bandwidths could not be discriminated at the higher common edge frequencies where some of the wider bandwidths were discriminable. This could be because the cues for discrimination were closer together for the narrow bandwidth stimuli or it could be that the subjects used the lower edge cues for the wider bandwidth stimuli. Of course if the noise band fell within a region of hearing impairment then this could limit the bandwidth of the higher noise band, creating additional cues for perception. This was seen for some of the subjects, with the wider bandwidths a reasonable level of performance was achieved at a common edge frequency that was the same as the highest frequency eliciting a response in audiometry. This means that even though the main part of the higher-frequency noise band fell in a region where there was little residual hearing, subjects could perform the task. Subjects were most likely using the low-frequency portion of the higher-frequency noise for discrimination. Examples of this are: subject BW for bandwidths of 250, 500 and 1000 Hz; subject MB for bandwidths of 250 and 500 Hz; subject YL for the 500-Hz bandwidth; subject TH for the 500-Hz bandwidth; and subject AV for the 250-Hz bandwidth.

The results with the 125-Hz wide noise suggest that the use of narrow band noises may be of value for listeners such as BB and MB, whose auditory areas are essentially limited to 500 Hz. The main problem in the use of noise bands as narrow as 125 Hz is that they have inherently high amplitude fluctuations, which may limit the extent to which they can convey amplitude modulation based on the speech amplitude envelope. However one of the advantages of the narrower noise bands is that they have more regular zero-crossings, making them easier to discriminate using temporal cues.

The use of wider bandwidth noises limits the number of bands within the auditory area, and is only likely to be effective for listeners such as JH, DH and BW, who have wider residual frequency ranges. This problem is illustrated by the decrease in the predicted number of discriminable noise bands as the bandwidth increases. However the centre frequencies of two edge-to-edge wide-band noises are farther apart than those of two narrow-band noises, so the use of wide noise bands would allow for a greater distinction between encoded stimuli.

For the narrower bandwidth noises there was a significant correlation between duration thresholds and hearing loss at the common edge frequency. This indicates that the cue used was nearer in frequency to the common edge frequency in the narrow band noises than the wider noises. It is probable that subjects were basing discrimination on the centre frequency of the noise bands, which for narrower noises is closer to the common edge frequency. This does suggest that hearing thresholds could be an indicator as to the ability the subject might have for discriminating the stimuli.

### 3.6 Summary and conclusions

The results of this experiment show that listeners with severe-to-profound hearing impairments retain sufficient hearing abilities to be able to discriminate a range of noise stimuli at durations comparable to those of voiceless fricatives in running speech. The noise stimuli varied in centre frequency and bandwidth and the duration was varied adaptively. All subjects were able to discriminate at least two spectrally adjacent 250-Hz wide noise bands from one another and two 500-Hz wide noise bands from one another.

The ability to perform the task was limited by degree of hearing loss and by the frequency range of hearing. However, the results showed that for the narrow noise bands, at least, if the absolute threshold was less than 80 dB then the subjects were typically able to discriminate the stimuli.

The results have positive implications for a strategy of encoding the voiceless excitation of obstruents using noise bands fitted to the individual's hearing loss because they show that this group of subjects can discriminate spectrally adjacent noise bands. This is obviously not the same as being able to identify the stimuli but it at least indicates that useful residual hearing abilities may still exist for listeners with severe-to-profound losses.

# CHAPTER FOUR

## 4 THE DISCRIMINATION OF THE BANDWIDTH OF NOISES BY NORMAL HEARING AND HEARING-IMPAIRED LISTENERS

### 4.1 Introduction

Bandwidth is one cue to fricative identity that could represent a potentially useful additional dimension in a simplified encoding scheme. The better the discrimination abilities of the listener, the more potential there is for flexibility in the encoding strategy, thus allowing the stimuli to be as similar as possible to the original sounds to improve naturalness and facilitate training.

The previous experiment showed that many severe-to-profoundly hearing impaired listeners retain the ability to discriminate to some degree, spectrally adjacent bands of noise of the same bandwidth (on a linear scale). The purpose of this experiment was to look at whether bandwidth could be a useable cue for speech perception by hearing impaired listeners. The experiment reported here examined the ability of severe-to-profoundly hearing-impaired subjects to discriminate the bandwidth of noises when the spectra of the two stimuli were centred at the same frequency. In a second stage of experimentation, work was carried out to determine if subjects were more reliant on the upper, lower or both edges of the noise bands for performing the task. This was achieved by fixing the upper or lower frequency edge of both stimuli. The two tasks will be explained separately in the following text.

### 4.2 Task 1. Bandwidth Discrimination with fixed centre frequency

This experiment examined the discrimination of noise bandwidth for noise bands of fixed centre frequency. The aim was to discover if severe-to-profoundly subjects retained any useful discrimination abilities that could be utilised for the identification of encoded fricatives.

#### *4.2.1 Subjects*

Five hearing-impaired and four normally hearing subjects took part. They all had previous experience in psycho-acoustic and speech perceptual experiments. The hearing-impaired subjects were aged between 35 and 75 and had post-lingual severe-to-profound sensorineural hearing losses. The normally hearing group were aged between 25 and 45. All subjects underwent hearing tests to ensure the reported loss was correct. Audiograms for the hearing-impaired subjects are shown in Figure 1.

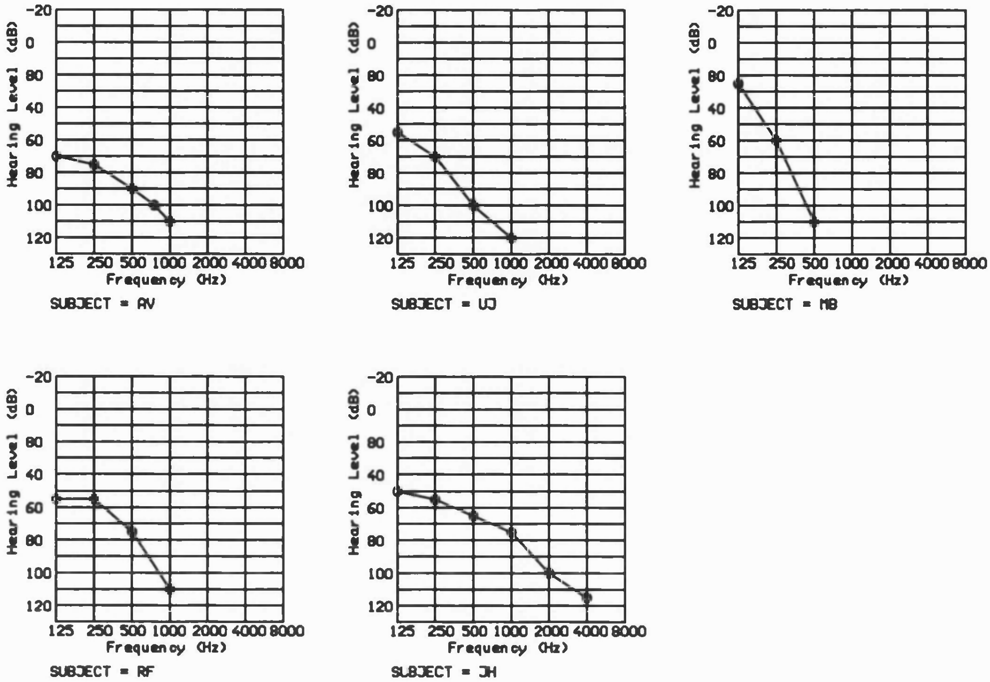



Figure 1. Audiograms for the hearing-impaired subjects

### *4.2.2 Task and conditions*

Subjects were required to discriminate noises of different bandwidths with a fixed centre frequency. The stimuli were band-pass filtered noises centred at 300 Hz with duration of 80 ms. A low centre frequency was selected to ensure that the stimuli were audible to all subjects. It was assumed that if subjects could discriminate the stimuli at short durations they could probably discriminate them under easier conditions when the durations of the stimuli were longer. Bandwidths of the noises ranged from 124 to 500 Hz. 500 Hz was chosen for the widest bandwidth because it was wide enough for there to be a large difference between the narrow and wide band stimuli but not so wide that the majority of the information fell outside the hearing range of some subjects. The narrowest bandwidth was chosen to be 124 Hz because this was wide enough to be useful as a bandwidth for encoding fricatives but not so narrow as to give difficulties with audibility; when the noise bandwidth is too narrow, subjects with narrow dynamic ranges have a tendency to choose a lower listening level to avoid discomfort due to the strong inherent fluctuations. This can make the stimuli hard to discriminate because much of the sound falls below absolute threshold, and could also lead to audible loudness fluctuations being used as a cue to discrimination. For the purpose of this experiment we tried to eliminate loudness cues so we could assess how well subjects could discriminate stimuli without the use of loudness cues. However for the purpose of an encoding strategy, loudness cues might help to improve identification.

There were two main conditions. In condition 1, the fixed reference stimulus had the broader bandwidth; this being 500 Hz, and the narrower bandwidth was varied. In condition 2 the reference stimulus had the narrower bandwidth, which was fixed at 124 Hz, and the broader bandwidth was varied.

One of the hearing-impaired subjects (MB) only participated in condition 1, because of work pressures and one normal-hearing subject (JD) moved away so could only complete task 1 and did not perform the full battery of tests described in task 2.

### *4.2.3 Stimulus generation and control*

The noise stimuli were derived from a white-noise source, and band-pass filtered using two Kemo VBF/8 filters giving an overall slope of 90 dB/octave. Each stimulus was gated on and off with a half cycle of a raised-cosine function lasting 5 ms. A Masscomp 5400 computer was used to control stimulus presentation. The stimuli were gated by multiplying the noise with a gating signal derived from a 12-bit D/A converter. After gating and filtering, the stimuli were sent along a balanced line to a sound-attenuating chamber. A Yamaha amplifier, Hatfield manual attenuators, Charybdis programmable attenuators and a potentiometer wheel were used to control the output levels of the stimuli. The stimuli were balanced in loudness and jittered in level as described below.

Stimuli were presented monaurally through Beyer Dynamic DT48 headphones; electronic equalization was provided digitally via a Barr and Stroud EF8 programmable filter to give a flat headphone frequency response down to about 70 Hz. A microphone mounted on the headphone grille and an FFT analyser (Ono-Sokki CF-910) were used to monitor the signal levels and spectra of the stimuli.

### *4.2.4 Procedure*

Psychometric functions were measured in conditions 1 and 2 for a range of bandwidth ratios, using a three-interval two-alternative forced-choice paradigm. In condition 1, the 500-Hz wide noise was presented in the first interval and the narrower bandwidth noise was presented in either the second or third interval with a 500-Hz wide noise in the other interval. In condition 2, the 124-Hz wide noise band was played in the first interval and the wider bandwidth noise was presented in either the second or third interval, with a 124-Hz wide noise in the other interval. The subjects had to decide whether the second or third stimulus was the odd one out and press the appropriate button on a response box. Feedback lights were used after each trial to indicate the correct response. Within a run, the bandwidths of the two noises were fixed and a percent correct score was obtained from a hundred trials. Each point on the psychometric function was repeated in a different session and the scores from the two sessions were combined producing a percent correct score from 200 presentations. All subjects received a minimum of 4 hours training before data collection began. At the start of each session, the threshold for detecting the noise bands was checked to ensure that the subject's thresholds had not changed. If the threshold for any of the noise bands was more than 10 dB different from the previous session then a further testing session was carried out with re-balanced loudness and the results from the two sessions where the thresholds were most similar were used.

#### 4.2.4.1 Loudness balancing

Three precautions were taken to ensure that discrimination was not based on loudness differences. For the hearing-impaired subjects, all stimuli were presented through a filter whose characteristics were based on equal-loudness judgements for pure tones (see below for details). For all subjects, the two stimuli used in a given run were additionally matched in loudness. Finally, a random amplitude jitter was applied to stimuli within each trial, as described below.

#### 4.2.4.2 Equal-loudness frequency shaping

Prior to testing, an individually tailored equal-loudness FIR (finite impulse response) filter was set up for each hearing-impaired subject. The filter parameters were individually calculated from a loudness balancing experiment. The procedure involved subjects balancing the loudness of 200-ms tones with 5-ms half-cycle raised-cosine ramps at 125-Hz intervals from 125-1125 Hz (where possible). Balancing began by comparing the loudness of a 500-Hz tone at comfortable listening level with the loudness of a 625-Hz tone. The two tones were played in continuous alternation and the level of the 625-Hz tone was adjusted until it appeared equally loud to the 500-Hz tone. This procedure was repeated for all the audible tones within the 125 - 1125 Hz range above and below 500 Hz. 500 Hz was always used as the frequency of the reference tone. Each match was made three times and the average level of the matches was used. The attenuation values required to produce the average levels were recorded at all frequencies and a filter shape interpolated. The values obtained in this way were used to define a FIR filter implemented on a Barr and Stroud EF8 programmable filter.

#### 4.2.4.3 Pair-wise loudness matching of stimuli

At the start of each run, subjects were required to rotate a potentiometer wheel whilst hearing the two noises in alternation. The potentiometer controlled the level of one noise while the level of the other noise was fixed. The subjects were instructed to adjust the wheel to find the point at which the two sounds appeared to be equally loud. They were advised to pass through the equal loudness point a few times to ensure accurate balancing. This procedure was carried out three times and the mean level at the matching point was used as the point of equal loudness. Once the equal loudness point was determined the subject listened to the stimuli again to ensure they were comfortably loud. If necessary the output level was adjusted.

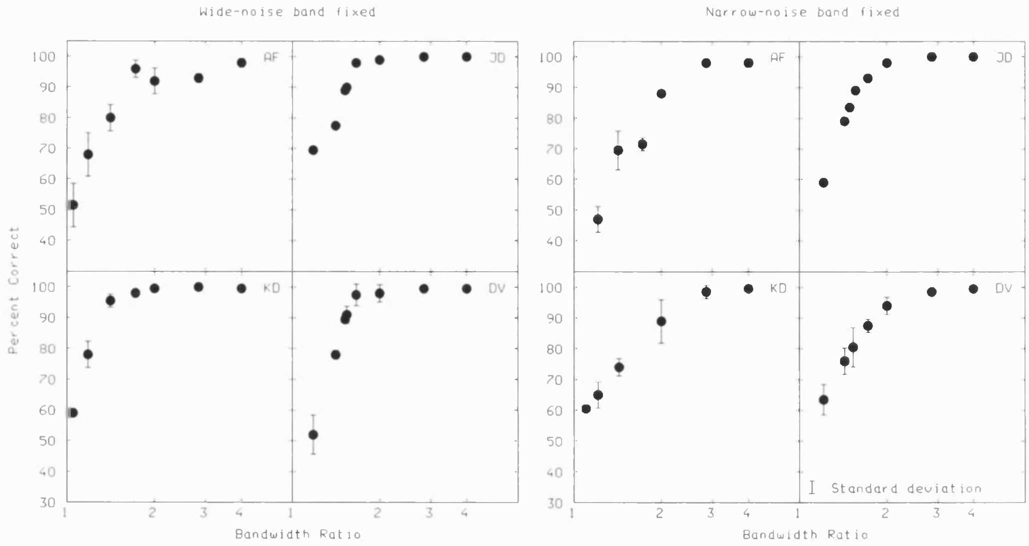
#### 4.2.4.4 Level Jitter

In order to eliminate any residual loudness cues due to balancing errors, levels were randomly varied over a 3-dB range from stimulus to stimulus for the hearing-impaired group and over a 6-dB range for the normally hearing subjects. The level of jitter was determined from a pilot study to assess the amount of jitter which was sufficient to eliminate loudness cues occurring due to errors in balancing but not so large as to have a detrimental effect on performance. The normally hearing subjects were able to tolerate a higher level of jitter than the hearing-impaired subjects.

It is most likely the case that the hearing impaired subjects could not tolerate such large jitter levels because loudness recruitment caused the fluctuations in the noise and the overall level differences to become exaggerated (Moore et al, 1996).

### *4.2.5 Results of task 1*

Figure 2 shows psychometric functions from conditions 1 (left panels) and 2 (right panels) for the normally hearing subjects. Figures 3 and 4 show data for the hearing-impaired subjects for conditions 1 and 2, respectively. The percent correct is plotted as a function of the bandwidth ratio. The vertical lines represent plus and minus one standard deviation of the results across sessions. The mean variance for the repeatability of the data, i.e. from the first session to the second session of data collection was: 36.75 and 47.75 for the hearing-impaired group in the wide band and narrow bands respectively,




and 15.45 and 11.76 for the wide band and narrow band conditions respectively for the normal hearing group.

Figure 2. Psychometric functions for the normal hearing group for condition 1 wide-band noise fixed (left) and condition 2, narrow-band noise fixed (right). Each panel shows the results from an individual subject. The percent correct score is plotted as a function of the bandwidth ratio.

The functions for the normal-hearing listeners are fairly smooth for both conditions. The slopes after the knee point in the psychometric functions tend to be fairly sharp particularly for condition 1. The slopes for condition 2 tend to be slightly shallower. The knee points for condition 1 tend to be at a smaller bandwidth ratio than for condition 2. The bandwidth ratios at the knee point tend to be less than 2 in condition 1 and are typically between 2 and 3 for condition 2.

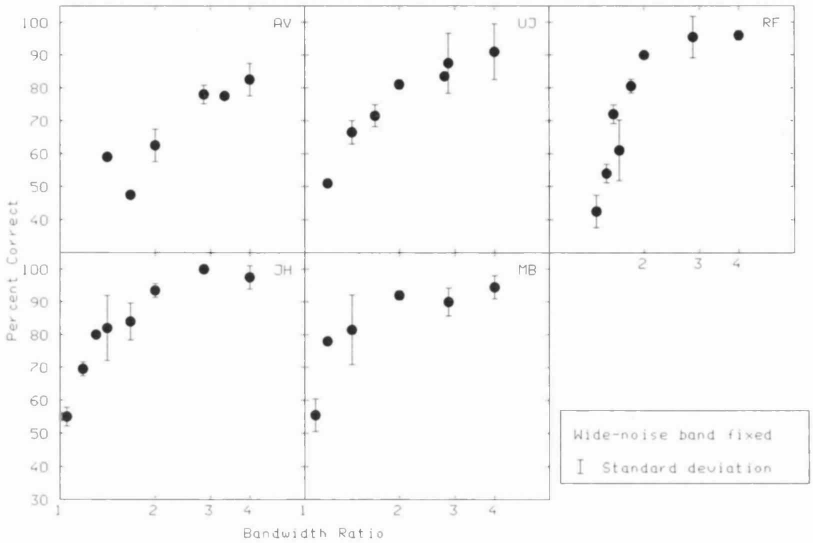



Figure 3. Psychometric functions for the hearing-impaired group in condition 1. Each panel shows the results from an individual subject. The percent correct score is plotted as a function of the bandwidth ratio.

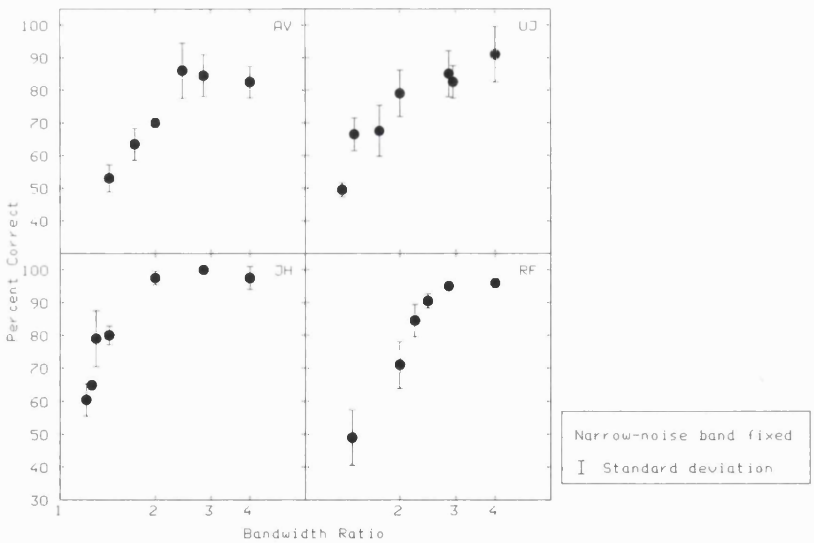



Figure 4. Psychometric functions for the hearing-impaired group in condition 2. Each panel shows the results from an individual subject. The percent correct score is plotted as a function of the bandwidth ratio.

The psychometric functions shown in figures 3 and 4 for the hearing-impaired group are also reasonably smooth but there is more variability between subjects than was the case for the normal-hearing subjects. The standard deviations are also typically larger than those obtained for the normal-hearing subjects. Subject AV performed most poorly of all the hearing-impaired subjects, and never achieved 100% correct under any of the conditions. Subject UJ also found the task very difficult and only occasionally scored 100%. The other subjects performed in a more systematic way.

The psychometric functions (percent correct as a function of the logarithm of the bandwidth ratio) were fitted using probit analysis (Finney, 1971) and the 75% point estimated. The 75% threshold point was used as a way of comparing performance across subjects and between the normally hearing and hearing-impaired groups. The 75% level was chosen because this point represented a bandwidth ratio where performance was above chance (50 % for a two-alternative forced-choice task) but well below the ceiling level.

Figure 5 shows the estimated 75 % points for all the subjects. The initials under each point refer to a subject. The open squares and diamonds indicate the 75% point estimated from conditions 1 and 2, respectively.

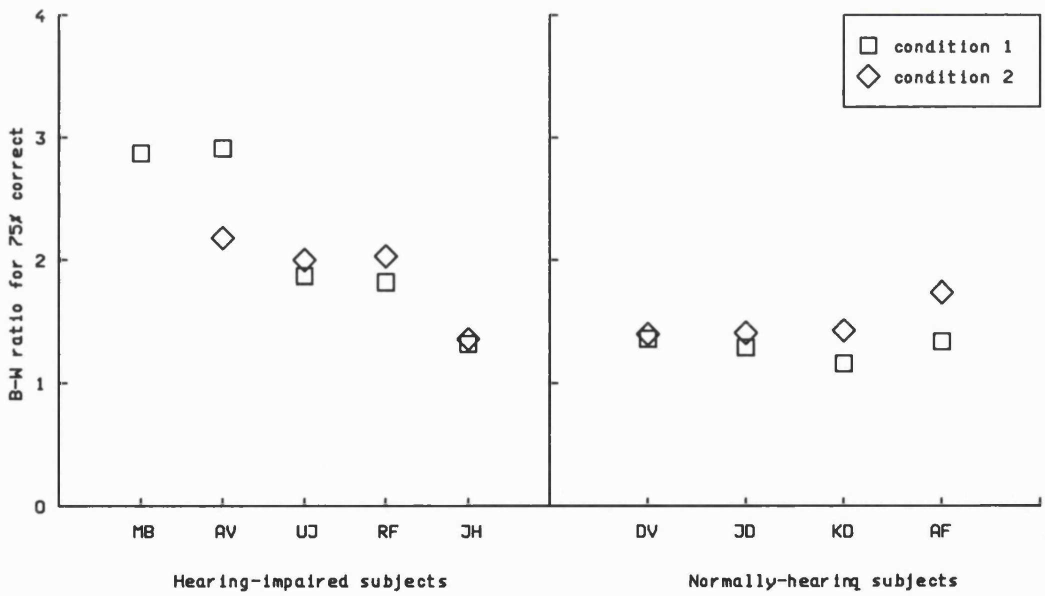



Figure 5. Estimated 75% points for all subjects in conditions 1 and 2. The results for the hearing-impaired and normal hearing subjects are shown in the left and right panels,

respectively. The squares and diamonds show the 75% estimated values for condition 1 and 2, respectively.

All subjects showed an ability to discriminate the stimuli to some extent. For the normally hearing group, the mean threshold (75% point) bandwidth ratios were 1.30 and 1.50 for conditions 1 and 2, respectively. The threshold ratio of 1.30 for condition 1 indicates that bandwidths of 500 Hz and 385 Hz were discriminable, an absolute difference of 57.5 Hz on each edge, which when translated into percentage terms is 12% on the upper edge and 115% on the lower edge. At threshold the lowest frequency edge was used as the reference frequency for calculation of the percentage differences, this happened on both the lower and upper edges of the noise bands. This was the case regardless of which noise was fixed to allow comparisons to be made. The edge frequency difference on either the upper or lower edge was divided by the reference frequency and multiplied by 100. This was the approach used by Pickett et al. (1965) to calculate the percentage differences. A ratio of 1.50 in condition 2 indicates that bandwidths of 124 Hz and 186 Hz were discriminable. This corresponds to a difference of 31 Hz on each edge, a percentage difference of 9% on the upper edge and 15% on the lower edge.

For the hearing-impaired group, the mean threshold bandwidth ratios for the three best subjects, UJ, RF and JH, were 1.67 and 1.80 for conditions 1 and 2, respectively. Subjects MB and AV had much larger threshold bandwidth ratios in condition 1 than subjects UJ, RF and JH. If the results from all the hearing-impaired subjects are used to calculate a mean bandwidth ratio at threshold, then the result is 2.16 for condition 1 and 1.89 for condition 2. A threshold ratio of 1.67 indicates that bandwidths of 500 and 299 Hz were discriminable. This corresponds to an edge difference on both the upper and lower edges of 100.5 Hz, which when translated in to a percentage difference (of the cut-off frequency of the fixed noise), is 22% on the upper edge and 201% on the lower edge. The 1.80 threshold ratio for condition 2 indicates that bandwidths of 124 Hz and 223 Hz could be discriminated, a difference of 49.5 Hz on each edge, and a percentage difference of 14% on the high-frequency edge and 26% on the low- frequency edge.

### 4.3 Task 2. Discrimination of spectra with single-edge frequency cues

A second task was run to assess if the subjects were using any particular edge of the noise bands for discriminating the stimuli. For the hearing-impaired subjects in particular, where hearing ability is generally better at lower frequencies, it was thought possible that only the lower edge frequencies were being used as a discrimination cue. Hence, stimuli were generated with either the upper or lower edge frequencies of the stimulus pair set to the same frequency, the other edge frequencies differing.

It is important to note that by looking at the effectiveness of edge cues in this way it also changes the centre frequency of the noise bands. It was thought that it would be interesting to see if the listeners were affected by this additional cue and in some cases it could be possible to tease out which cues were being used to perform the task.

#### *4.3.1 Procedure*

Stimuli with the bandwidth ratios corresponding to the 75 % point estimated in conditions 1 (wide band fixed) and 2 (narrow band fixed) of task 1 were used to assess if subjects were favouring a particular edge of the stimuli for discrimination. The first stage of testing was to check the accuracy of the 75% point for both conditions 1 and 2 by repeating the presentation of the two noises estimated by the 75% point and determining the percent correct score. This was to ensure that the estimated 75% point was actually realistic. For each set of two noises, one set for condition 1 and one for condition 2, there were two further testing stages in which either the upper or lower edge frequency of the variable stimulus was fixed at the same frequency as for the reference noise band. For condition 1 the edge frequency was 50 Hz for the lower edge and 550 Hz for the upper edge. For condition 2, the edge frequency was 238 Hz for the lower edge and 362 Hz for the upper edge.

The other edge of the variable stimulus was set to the same frequency as the corresponding edge in the 75% estimate. A stylised representation of an example set of stimuli is shown in figure 6. The example is based on a 2:1 bandwidth ratio at the 75% point.

Scores for each condition were based on 200 three-interval trials, 100 run in one session and a 100 run on a separate day. In all other respects the procedure was the same as that for task 1, as were the subjects.

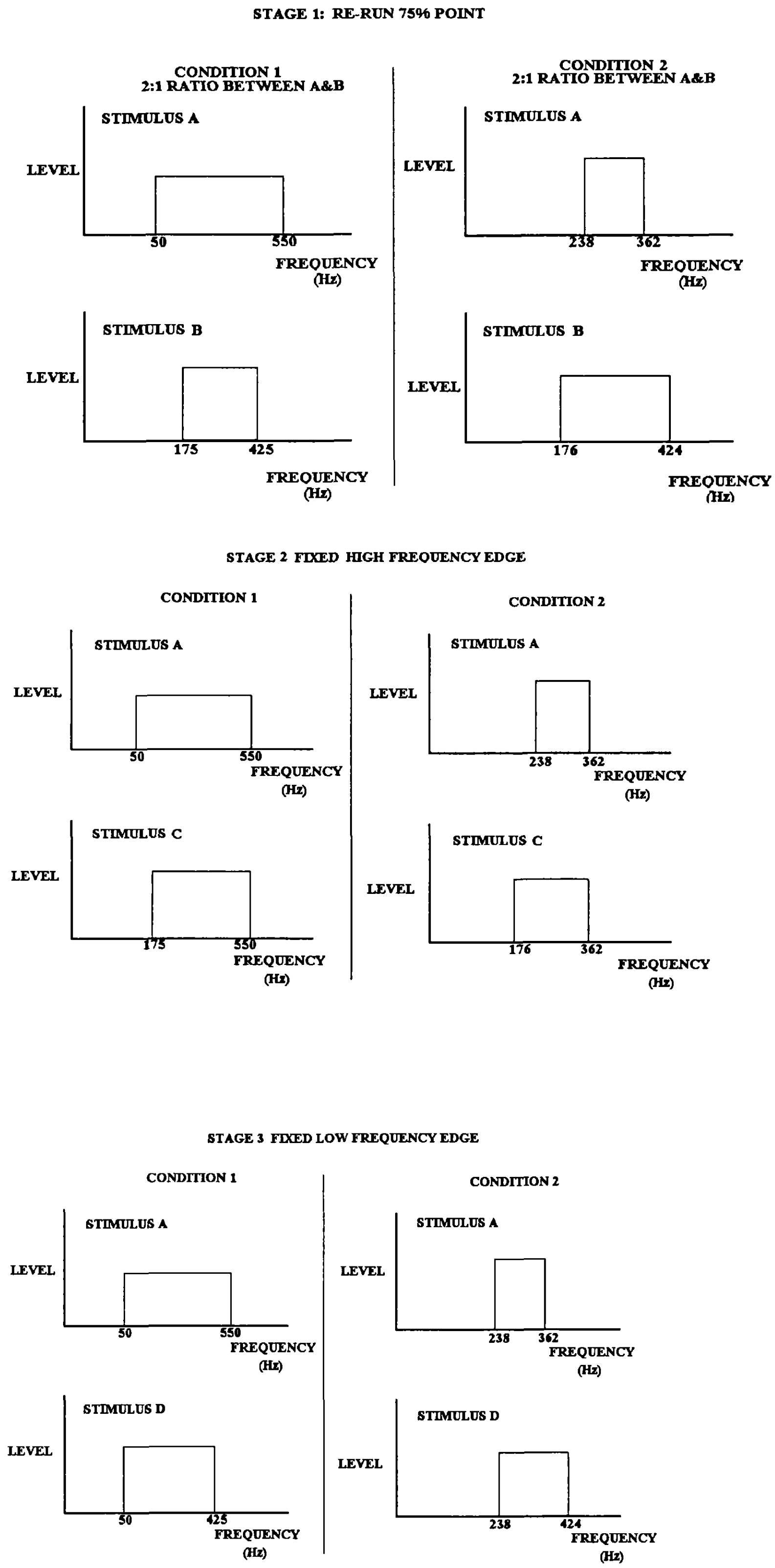



Figure 6. Stylised diagram of a set of stimuli in task 2.

### 4.3.2 *Results*

Table 1 shows the results for the both the hearing-impaired and normal- hearing subjects. The estimated 75 % correct bandwidth ratios estimated from the Probit analysis carried out on the psychometric functions from conditions 1 and 2 in task 1 are shown. The table also shows the percent correct score obtained when these conditions were re-run. The scores for the repeated 75% points ranged from 67-83.5%. This good repeatability indicates that the estimates from the probit analysis were reasonably accurate. The scores are also shown for the conditions with either the upper or lower edges fixed. Note that in table 1 some of the points for subjects MB and JD were not collected because the subjects could not complete all of tests.

**Hearing-Impaired group**

| | Condition 1 (wide-band fixed) | | | | Condition 2 (narrow-band fixed) | | | |
|---|---|---|---|---|---|---|---|---|
| Subj | Bandwidth ratio at the 75% point | Percent correct score for 75% point re-test | Percent correct score with lower edge variable | Percent correct score with upper edge variable | Bandwidth ratio at the 75% point | Percent correct score for 75% point re-test | Percent correct score with lower edge variable | Percent correct score with upper edge variable |
| UJ | 1.87 | 69.0 | 51.0 | 56.0 | 2.00 | 69.0 | 69.0 | 55.0 |
| JH | 1.32 | 72.0 | 63.0 | 70.0 | 1.36 | 74.5 | 70.0 | 72.0 |
| RF | 1.82 | 83.0 | 74.0 | 64.0 | 2.03 | 70.0 | 72.5 | 68.0 |
| AV | 2.91 | 69.0 | 71.5 | 46.5 | 2.18 | 74.0 | 54.5 | 65.0 |
| MB | 2.87 | 67.0 | | | | | | |

**Normally hearing group**

| | Condition 1 (wide-band fixed) | | | | Condition 2 (narrow-band fixed) | | | |
|---|---|---|---|---|---|---|---|---|
| Subj | Bandwidth ratio at the 75% point | Percent correct score for 75% point re-test | Percent correct score with lower edge variable | Percent correct score with upper edge variable | Bandwidth ratio at the 75% point | Percent correct score for 75% point re-test | Percent correct score with lower edge variable | Percent correct score with upper edge variable |
| DV | 1.36 | 83.5 | 68.0 | 84.0 | 1.40 | 71.0 | 60.0 | 60.5 |
| JD | 1.29 | 80.5 | | | 1.41 | 76.5 | | |
| KD | 1.16 | 71.0 | 66.0 | 77.0 | 1.43 | 70.5 | 63.5 | 67.5 |
| AF | 1.34 | 76.5 | 69.0 | 82.5 | 1.74 | 72.0 | 85.0 | 67.0 |

Table 1. The results of task 2, conditions 1 and 2. 75 % thresholds estimated from probit analysis are shown along with the results from re-running that point and also for the conditions where the upper or lower edge of the noise was fixed.

Figures 7 and 8 show the results of task 2 for the normal-hearing and hearing-impaired subjects, respectively. The empty diamonds show scores obtained when the predicted 75 % points were re-run. The upward-pointing arrows show the scores when the upper edges of the noises were fixed and the downwards-pointing arrows show the scores when the lower edges were fixed.

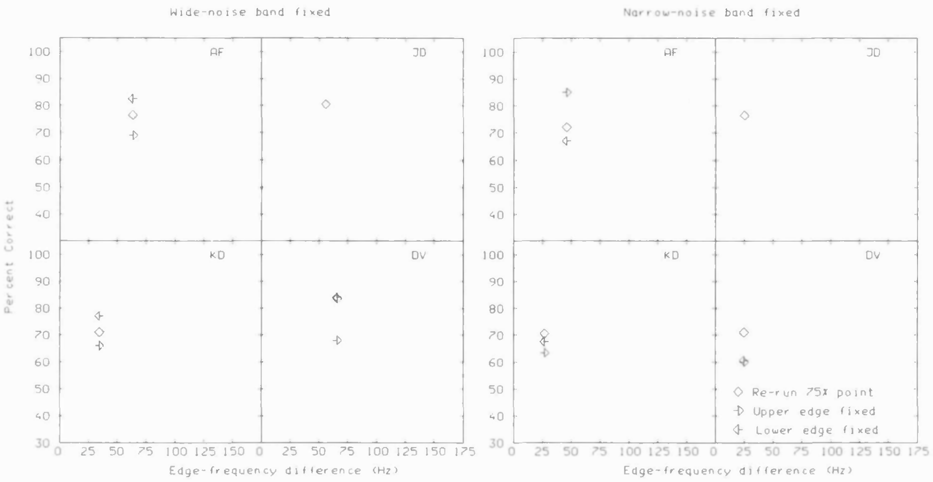



Figure 7. The results from task 2 for the normal-hearing subjects

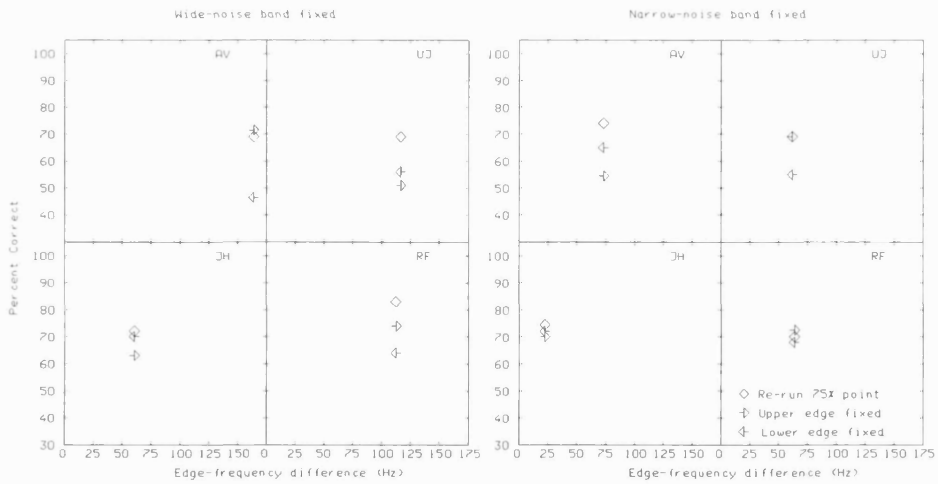



Figure 8. The results from task 2 for the hearing-impaired subjects

It is interesting to note that seldom is performance much better with both edges varying than it is with just one edge varying. Consider the results from task 2 using the wider bandwidth reference stimulus from the normal-hearing subjects and the hearing-impaired subject JH. Performance was at the same level or even better when only the high-frequency edges of the stimuli were varied than when both edges were varied. This could indicate that these listeners' performance was determined only by the high-frequency edge cues when both edges were varying. However, particularly in the cases where performance is better with one edge than two, subjects could have potentially used the centre frequency cue for discrimination in task 2. For the hearing-impaired subjects whose high-frequency hearing was not as good as that of JH, the results are slightly different. Subjects RF and UJ appeared to use both edges for discrimination, because neither of the asymmetric conditions on their own gave a similar level of performance to that when both edges were available. Subject AV, who had the poorest hearing, probably relied totally on the low-frequency edge of the stimuli. It is also possible that in task 2 subjects could have been confused by the additional centre frequency cue. The results from task 2 using the narrower bandwidth reference stimulus show that the normal hearing subjects DV and the hearing-impaired subject AV appeared to use both edges for discrimination in the symmetric condition. This is indicated by the fact that variation in either edge alone did not produce the same level of performance as when both edges were varied. Probably the subjects were integrating cues from both band edge regions. Subject UJ performed at the same level with the low frequency edge varied as for both edges varied, and performed much worse when the high frequency edge was varied. This would suggest that UJ could only take advantage of cues from the low-frequency edge. The normal-hearing subjects AF and KD and the hearing-impaired subjects JH and RF performed at the same level or better in both of the one edge varied conditions as for both edges varied. This could indicate that subjects could have used either edge in the two edge varied condition or it could suggest that subjects were using the centre frequency cue to perform the task in the asymmetric condition.

### 4.4 Discussion

Speech is a complex and dynamic signal, continuously varying in frequency, amplitude and time. The production of one phoneme can vary dramatically with such factors as the preceding and following speech sounds and the rapidity of speech. Although this work appears to treat speech as a very static entity focusing on specific speech cues, the final aim is to provide an encoded signal that varies in a similar way to the original speech but with some of the features simplified to improve perception for those people who cannot cope with the complexity of the original speech. Such an encoding strategy would include not only bandwidth but also other cues, for example intensity and centre frequency. Bandwidth is expected to play an important role in improving naturalness and providing a further cue to improve perception.

The results of the experiments described in this chapter have been encouraging. All the subjects tested could, to some degree, perform the tasks. This suggests that even profoundly hearing-impaired subjects have hearing abilities in the low-frequency region that potentially could be used to identify encoded speech stimuli using the dimension of bandwidth. These abilities are also very relevant to the use of other strategies such as frequency transposition (Velmans, 1973; Braida et al, 1979, Rosenhouse, 1989).

In general, the hearing-impaired group performed more poorly than the normally hearing group, except for JH whose results fell within the normal-hearing range. For both groups, 75% bandwidth ratio thresholds were usually slightly higher when the narrower bandwidth was fixed; the results for the impaired listener AV were the only exception. The thresholds could have been higher for the narrow bandwidth fixed because for a given bandwidth ratio, the change in edge frequency is greater for the wide band fixed than for the narrow band fixed. For example, for a bandwidth ratio of 2:1, a noise of 500-Hz bandwidth was discriminated from a noise of 250-Hz bandwidth in condition 1 (edge frequency differences of 125 Hz), while a noise of 124-Hz bandwidth was discriminated from a noise of 248-Hz bandwidth in condition 2 (edge frequency differences of 62 Hz).

It is possible that the severe-to-profound hearing-impaired listeners with greatly reduced frequency selectivity could have relied heavily on temporal cues to perform the task. Possibly some of the wider bandwidth stimuli fell within more than one auditory filter, but it is most likely that the narrower stimuli would have fallen into one auditory filter. The results would suggest that subjects UJ, JH and RF could have been using across-filter cues for the wide-band fixed condition because they performed relatively well compared to condition 2. However subjects AV and MB performed poorly in condition 1 compared to condition 2, showing that this condition was more difficult and possibly if they were relying on within-filter cues, these would be much weaker for the wider bandwidth stimuli.

In support of the finding of Pickett et al. (1965), the performance of the normal-hearing and hearing-impaired subjects were rather similar, given the degree of hearing loss of some of the subjects. Threshold bandwidth ratios were 1.30 and 1.67 in condition 1 for the normal-hearing and hearing-impaired subjects respectively, and 1.50 and 1.80 in condition 2. It is possible that the similarity between the normal hearing and hearing-impaired subjects in this experiment was due to discrimination being based mainly upon temporal information. It appears that bandwidth could be a very important cue for encoding fricatives because of its accessibility to hearing-impaired subjects.

### 4.4.1 *Is discrimination based on the upper, lower or both edges?*

In task 2, performance was assessed with either the upper or lower edge frequency fixed, so that the edge cues from only one edge were varied. This was intended to give some insight in to whether performance was based upon attending to one edge, either edge or both edges. To perform the task with just one edge varying subjects were possibly attending to differences in the edges of the noises, or a difference in the centre frequency of the two noises. Whereas when both edges were varied and the centre of the noise bands was fixed the edge cues were the only ones available.

The noise bands used in this experiment are similar to those used by Fastl (1971). He found that both edges of a 600-Hz wide noise band in the low-frequency region elicited a strong pitch sensation for normally hearing listeners. At higher centre frequencies, the pitch sensation related more to the centre frequency of the band of noise and not the edges. The stimuli used here were low in frequency, so it is possible that edge pitch cues were available. The stimuli used had fairly sharp cutoff slopes (90 dB/Oct.) which increases the pitch strength of the edge pitch cues (Fastl, 1980).

### 4.4.2 *Outcome of task 2 using the wider bandwidth stimuli.*

Where the fixed reference stimulus had the wider 500-Hz bandwidth, the results from task 2 appear to be fairly straightforward. If the subject had normal hearing or useable hearing around 500 Hz, then the high-frequency edge varied condition gave higher scores than the low-frequency edge varied condition and the same or better score than both edges varied. This could indicate that listeners mainly used the high frequency edge for discrimination when both edges were available or that subjects were using centre frequency cues to perform the task when only one edge was available. It is possible that the low frequency edge alone was not as easy to use as the high frequency edge because the low-frequency edge would have fallen in a frequency region (50 - 100 Hz) where discrimination abilities are generally poorer (Harris, 1952) than at higher frequencies. The hearing-impaired subjects, who did not have such good hearing at 500

Hz, could not make such effective use of the high-frequency edges of the stimuli for discrimination. Subjects UJ and RF used both edges of the stimuli, whereas subject AV relied totally on the low frequency edge, which may explain why her ratio for the symmetric condition was typically much worse than for the other subjects.

### *4.4.3 Outcome of task 2 using the narrower bandwidth stimuli.*

Where the fixed reference stimulus had the narrower 124 Hz bandwidth, the results from task 2 were not so easy to interpret. This is probably because there were more perceptual cues available to the subjects. The low-frequency edge fell in a higher, more useable, frequency region, approximately around 180-240 Hz. Furthermore, the subjects who were not able to use the high frequency edge in condition 1 probably could in condition 2 because it was much lower, approximately around 360-420 Hz.

Subjects AV and DV appeared to use both edges in the two edge varying condition because neither edge on its own gave performance of the same level as both edges varying. Subject UJ performed better with the low-frequency edge varying alone than the high-frequency edge varying alone and scored at the same level as both edges varying which probably suggests that she was more able to use the cues from the low-frequency edge. All the other subjects performed at the same level or better in both of the one edge varying conditions than for the two edge varying condition. This could imply that subjects could use either edge to do the two-edge varied condition or that they were using centre frequency cues in the asymmetric condition.

The results from this study are consistent with those of Pickett et al. (1965). They tested subjects with severe cochlear damage and subjects with normal hearing. Discrimination of the cutoff frequency of lowpass filtered noise was carried out at a range of cutoff frequencies. They found a just-noticeable difference (jnd) at a reference cutoff frequency of 250 Hz in normally hearing subjects of 12%, which represents a 30 Hz difference. At the higher frequencies their jnds ranged from 3 to 5%. The jnd for the hearing-impaired subjects in the Pickett et al. study was 15% at 250 Hz, which corresponds to an absolute frequency difference of 37.5 Hz. Performance worsened at higher cut-off frequencies.

The normal-hearing subjects and subject JH performed better with the high-frequency than the low-frequency edge varied for the wide band fixed condition and at the same level approximately as for both edges varied. This could potentially mean that they attended to the high-frequency edge when doing the two-edge discrimination task. If we assume that these subjects were not attending to the low-frequency edge, we can compare our results to those of Pickett et al. (1965). The mean 75% bandwidth ratio in condition 1 for the normal-hearing subjects and JH was 1.30, corresponding to a frequency difference of 57.5 Hz on the high-frequency edge a percentage difference of 11.7%. This is similar to the 12% value found by Pickett et al. The cut-off frequency in our experiment was slightly higher (492.5 Hz) so it would perhaps be predicted from their results that the jnd in our experiment would be slightly less. Nevertheless our results are certainly in a similar range to those of Pickett et al. indicating that normal-hearing subjects are relatively insensitive to spectral differences in noise stimuli.

### 4.5 Summary

These results show that severe-to-profoundly hearing-impaired subjects were able to detect changes in the bandwidth of noise stimuli under conditions where the loudness of the bands did not provide useable discrimination cues. This ability could potentially be exploited in the encoding of phonetically relevant acoustic cues in aperiodic speech sounds. It remains to be determined whether bandwidth can be used as a cue for absolute identification as opposed to discrimination.

# CHAPTER FIVE

## 5 THE LABELLING OF NOISE BANDS AS FRICATIVES WITHIN A SPEECH-LIKE ENVIRONMENT BY HEARING-IMPAIRED LISTENERS

### *5.1 Introduction*

The work described in this chapter aimed to build on the information obtained from the previous experiments described in this thesis. It was previously determined that severe-to-profoundly hearing-impaired listeners often retain sufficient residual auditory abilities to discriminate between noise stimuli changing in bandwidth and also centre frequency. The discrimination abilities observed were sufficiently good to indicate that a selection of noise stimuli could be used to represent a range of speech stimuli. The experiment described in this chapter was set up to investigate whether a small subset of speech stimuli could be represented with a selection of acoustically simple noise stimuli in an identification task with severe-to-profoundly hearing impaired listeners.

The psychoacoustic investigations described in Chapters 3 and 4 showed that severe-to-profoundly hearing-impaired listeners often retain potentially useful residual auditory abilities for discriminating bands of noise. The subjects tested could all discriminate spectrally adjacent noise bands to some extent and also showed some ability to discriminate the bandwidth of noise bands centred on the same frequency. The study described in Chapter 2 showed that, for normal-hearing listeners, the centre frequency, bandwidth, and intensity of simple noise bands could be used to carry some of the acoustic cues associated with fricative perception. Here, similar cues, in the low-frequency range, are used to represent fricatives for the impaired listeners.

The questions that the current experiment aimed to address were:

1. Can the major spectral characteristics of voiceless fricatives be encoded to effectively represent place of articulation by the use of spectrally shaped noises mapped into the low-frequency hearing range of hearing-impaired listeners?
2. What might be an appropriate form of mapping for voiceless speech information, to make these cues optimally usable for hearing-impaired listeners?

Several different mappings were used which differed in complexity. Conventional frequency response shaping with amplification was used as a reference condition.

Speech processing was based upon a spectrographic analysis of the original speech tokens. The selected transformations of the original signal were designed to be achievable using real-time processing. Testing was carried out using auditory-alone, auditory-visual and visual-alone (lipreading) presentation modes. The main area of interest was the auditory alone condition. The auditory-visual condition was intended to facilitate adaptation to the auditory cues, while the visual condition served as a reference to indicate the contribution of auditory information in the auditory-visual condition.

## 5.2 Method

### 5.2.1 Stimuli

Stimuli were based upon natural speech tokens from an adult female talker with a Standard English RP accent. The original recordings were auditory-visual (A-V) tokens. On one audio channel was the speech signal and on the other the output from a laryngograph (Fourcin and Abberton 1977). There were 45 vowel-consonant-vowel (VCV) tokens in total, 15 tokens for each of three fricatives /f, s and ʃ/ in the context of 3 vowels /i, a and u/ e.g. /afa/, /ifi/, /ufu/. The second syllable of the VCV utterance was stressed.

The original A-V files were stored on a PC and the speech audio channel was transferred to a SUN-16 workstation for analysis and processing. By processing the original audio signals, it was possible to keep the durations of the processed stimuli temporally matched to the original audio signals, so that these could be readily synchronized with the video.

### *5.2.2 Spectrographic analysis*

The audio signals from all of the 45 tokens were spectrographically analysed. The spectrographic analysis was carried out using Entropics Waves+ software version 5.0. These analyses had an 8-kHz overall bandwidth, and were both wide-band (300 Hz) and narrow-band (45 Hz). The average spectrum of a 50-ms section from the centre of the steady-state portion of each fricative was measured. The chosen section was temporally centred in the noise-excited section. From this spectrum, measurements were made to determine the frequency limits within which most of the energy was concentrated and the frequency of the most prominent spectral peak. When there was more than one main spectral peak, the one with highest amplitude was chosen (for some of the /f/ utterances two peaks of similar level were measured, in these cases the higher frequency value was used because in the noise labelling experiment of chapter 2 people perceived the higher frequency noise bands as /f/). Bandwidth was taken as the frequency difference between the 3 dB down point below the lowest frequency of the main area of spectral energy and the 3 dB down point above the upper frequency limit of the highest energy region. Table I shows the results of these analyses.

Table I. A summary of the spectrographic analysis of the original fricative tokens.

| Fric. | Vowel | Low freq. edge (Hz) (LFE) | High freq. edge (Hz) (HFE) | Main peak (Hz) (MSP) | Mean for vowel context (Hz) LFE | HFE | MSP | Mean, collapsed across vowel (Hz) LFE | HFE | MSP |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1990 | 7550 | 7420 | | | | | | |
| | | 1990 | 7760 | 7690 | | | | | | |
| | /i/" | 1930 | 7600 | 7380 | 1998 | 7600 | 7464 | | | |
| | | 1960 | 7640 | 7480 | | | | | | |
| | | 2120 | 7450 | 7350 | | | | | | |
| | | 700 | 7650 | 7420 | | | | | | |
| | | 700 | 7600 | 7410 | | | | | | |
| /f/ | /a/ | 880 | 7740 | 7550 | 732 | 7634 | 7396 | 1283 | 7638 | 7401 |
| | | 930 | 7540 | 7200 | | | | | | |
| | | 450 | 7640 | 7400 | | | | | | |
| | | 1400 | 7590 | 7330 | | | | | | |
| | | 1000 | 7740 | 7500 | | | | | | |
| | /u/ | 965 | 7680 | 7500 | 1119 | 7680 | 7454 | | | |
| | | 1280 | 7670 | 7490 | | | | | | |
| | | 950 | 7720 | 7450 | | | | | | |
| | | 4750 | 7210 | 5440 | | | | | | |
| | | 4700 | 7530 | 5540 | | | | | | |
| | /i/ | 4800 | 7370 | 5360 | 4750 | 7396 | 5368 | | | |
| | | 4700 | 7440 | 5200 | | | | | | |
| | | 4800 | 7430 | 5300 | | | | | | |
| | | 4200 | 7480 | 5570 | | | | | | |
| | | 4090 | 7550 | 4850 | | | | | | |
| /s/ | /a/ | 4090 | 7800 | 5600 | 4122 | 7556 | 5146 | 4320 | 7461 | 5095 |
| | | 4060 | 7430 | 4750 | | | | | | |
| | | 4170 | 7520 | 4960 | | | | | | |
| | | 4180 | 7500 | 4300 | | | | | | |
| | | 4130 | 7510 | 4600 | | | | | | |
| | /u/ | 4150 | 7300 | 4550 | 4088 | 7430 | 4770 | | | |
| | | 4020 | 7450 | 4750 | | | | | | |
| | | 3960 | 7390 | 5650 | | | | | | |
| | | 1510 | 7200 | 3870 | | | | | | |
| | | 1730 | 7400 | 3900 | | | | | | |
| | /i/ | 1770 | 7350 | 3680 | 1832 | 7320 | 3822 | | | |
| | | 2090 | 7350 | 3790 | | | | | | |
| | | 2060 | 7300 | 3870 | | | | | | |
| | | 1800 | 7000 | 3745 | | | | | | |
| | | 1970 | 7500 | 3780 | | | | | | |
| /ʃ/ | /a/ | 1730 | 7400 | 3520 | 1832 | 7250 | 3626 | 1786 | 7267 | 3655 |
| | | 1990 | 7250 | 3435 | | | | | | |
| | | 1670 | 7100 | 3650 | | | | | | |
| | | 1860 | 7140 | 3600 | | | | | | |
| | | 1780 | 7200 | 3550 | | | | | | |
| | /u/ | 1460 | 7170 | 3530 | 1694 | 7232 | 3518 | | | |
| | | 1740 | 7400 | 3560 | | | | | | |
| | | 1630 | 7250 | 3350 | | | | | | |

The values for each of the three fricative categories were averaged across vowels and each of the five tokens to produce three mean sets of spectral measurements, one for each of the three fricative categories analysed. This obviously did not allow for the changes in spectra that occur for fricatives in different contexts, such as: changes in the mean spectrum level occurring for different utterances and individual talkers; dynamic changes occurring within the utterance of each fricative; and vowel-dependent dynamic spectral changes. This was intentional to allow the final stimuli to be fairly simple and to give the best possible opportunity for the processing to be implemented in a practical device.

There were some differences in the positions of the main spectral peaks and band edges for different vowel contexts but these were not taken into account for the purpose of the processing carried out here. The main differences were observed for utterances of /f/. The low frequency edge (LFE) using the /i/ vowel were typically higher in frequency than for the other two vowel contexts for the analyses with the /f/ phoneme. The /a/ vowel environment typically had the lowest LFE for the measurements taken with the /f/ phoneme. For /s/ and /ʃ/ phonemes, the LFE and MSP values in the /i/ environment were typically higher than for the other two vowels but there was not as much variability across vowel environments as was observed in the analyses with /f/. Also the vowel environments producing the lowest LFE and MSP values for the fricative were different for /s/ and /ʃ/, as opposed to /f/. For /f/, the lowest LFE and MSP values were seen using the /a/ environment whereas for /s/ and /ʃ/ the lowest LFE and MSP values were seen using the /u/ vowel context.

Appendix 1 shows spectrograms for different fricatives.

### *5.2.3 Processing*

Initially the vocalic sections of each token were analysed. The pitch was extracted for each vowel using the laryngograph waveform to determine the timing of the vocal fold closures occurring during the vowel utterance. The vowel was replaced by a sinusoid whose frequency matched that of the fundamental frequency of the vowel. This was done in order to remove formant structure and formant transitions. Formant transitions could be incongruent with the spectra of the synthesised frication noise and the purpose of the experiment was to see if the noise itself could carry the spectral information. Finally, a simple spectrum for vocalic sections maintains the spectral simplicity of the stimuli.

A noise was created for each token with exactly the same duration as the original frication noise. A new noise was created for each token to ensure that subjects did not use any specific feature of an individual noise for identification. The replacement noises had been processed in one of four ways listed below. Four sets of the 45 tokens were created, one for each of the processing conditions.

The conditions were:

***Flat*** **-** The average 3-dB bandwidth values for /f/, /s/ and /ʃ/ were used to specify the 3-dB bandwidth of the processed noise. The values for the LFE and HFE shown in Table 1 were divided by 8 to define values for the edge frequencies (LFE' and HFE') of a band-pass noise for each fricative. This shifted the spectral information of the stimuli into the audible frequency range of the subjects, while preserving relative bandwidth between stimuli.

***Shaped*** **-** These stimuli were designed to convey information about the main spectral peak. The average MSP value for each fricative was divided by 8 to define the spectral peak (MSP') for a low-frequency noise representing each fricative. The LFE' and HFE' values derived for the Flat stimuli were used as values for tapering the spectrum from the peak MSP' to the edges LFE' and HFE'. The source for the flat and shaped stimuli was the white noise output of an Ono-Sokki CF-910 FFT analyser. It was filtered through a Kemo 1/3 octave filterbank (Kemo system 2251 36 channel spectrum shaper). The resulting noises were digitised via a PC video and sound card (Videologic Mediaspace) and transferred to the SUN-16 Workstation.

***Natural*** **-** Stimuli were synthesized to create steady-state noises as similar as possible to the average representations of the original speech, within the same frequency range as the natural tokens. The noises for the Natural conditions were created using a Klatt synthesizer on the SUN-16 Workstation.

***Frequency compressed*** **-** The Natural stimuli were lowered in frequency by decreasing the sample rate by a factor of 8. A section of the down-sampled steady-state noise of the required duration was then selected. This process lowered the spectra of the stimuli to match the audible range of the listener whilst preserving the relative positions of the spectral peaks.

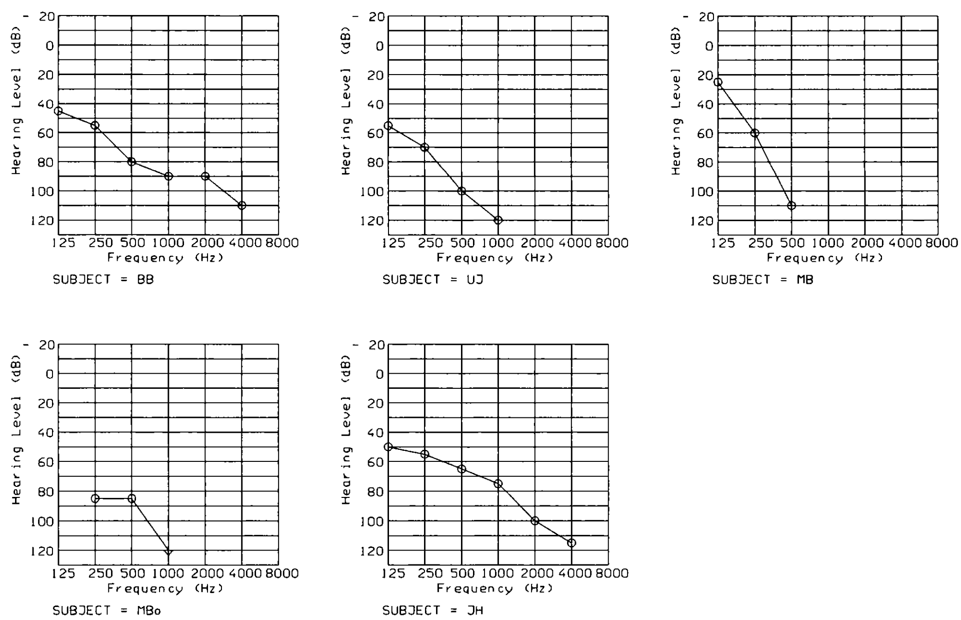
The noise for each token was inserted in between the two sinusoids that represented the vowels to create the final auditory signal. The rms (root mean squared) value of the noise for /f/ was 5 dB below that of the sinusoid and the value for the sibilants

/s/ and /ʃ/ was 5 dB above the level of the sinusoid. This allowed the sibilant/non-sibilant distinction to be cued by relative level, whilst ensuring that the frication always fitted within the dynamic range of the listeners. This happened for all processing conditions, including natural. Examples of the spectra of the stimuli are shown in Appendix 5.

### *5.2.4 Subjects*

Five severe-to-profoundly hearing-impaired subjects were tested. Their audiograms are shown in Figure 1. All subjects had previous experience of psychoacoustic and speech perceptual testing and were known to have cochlear hearing losses. All subjects had been involved in at least one of the noise spectrum discrimination tasks described in chapters 3 and 4. The losses were diagnosed as being cochlear on the basis of a lack of an air-bone gap on the audiogram, signs of loudness recruitment and normal tympanograms.

Figure 1. Audiograms of the subjects participating in the experiment.




### *5.2.5 Stimulus presentation*

Prior to testing, all subjects underwent an audiometric evaluation to determine their threshold of hearing, most comfortable listening level and highest comfortable level at frequencies ranging from 80 to 8000 Hz. Information gained from the evaluation was used to apply frequency response shaping according to an extended POGO formula. From 250 Hz upwards the frequency response rules followed those set out by

McCandless and Lyregaard (1983). The gain at 125 Hz was as specified by Faulkner et al. (1997). The POGO insertion gains at different frequencies as a function of hearing level are shown in Table II. The listener adjusted the overall gain for comfort. Care was taken to ensure that the stimulus delivery system remained linear for all subjects and stimuli. The levels of the stimuli being carefully controlled facilitated this and so therefore fitted within the relatively narrow dynamic ranges of the subjects, this would not be the case with natural speech.

| HL \ FREQ | 125 | 250 | 500 | 1000 | 2000 |
|---|---|---|---|---|---|
| 30 | 0 | 0 | 0 | 0 | 0 |
| 35 | 0 | 0 | 0 | 3 | 3 |
| 40 | 0 | 0 | 3 | 8 | 8 |
| 45 | 0 | 3 | 8 | 13 | 13 |
| 50 | 3 | 8 | 13 | 18 | 18 |
| 55 | 8 | 13 | 18 | 23 | 23 |
| 60 | 13 | 18 | 23 | 28 | 28 |
| 65 | 18 | 23 | 28 | 33 | 33 |
| 70 | 23 | 28 | 33 | 38 | 38 |
| 75 | 28 | 33 | 38 | 43 | 43 |
| 80 | 33 | 38 | 43 | 48 | 48 |
| 85 | 38 | 43 | 48 | 53 | 53 |
| 90 | 43 | 48 | 53 | 58 | 58 |
| 95 | 48 | 53 | 58 | 63 | 63 |
| 100 | . | 58 | 63 | 68 | 68 |
| 105 | . | 63 | 68 | 73 | 73 |
| 110 | . | 68 | 73 | 78 | 78 |
| 115 | . | . | 78 | 83 | 83 |
| 120 | . | . | 83 | 88 | 88 |
| 125 | . | . | . | 93 | 93 |

Table II. POGO insertion gain as a function of frequency and hearing level (HL)

Both hearing level measurements and testing were performed through the same set-up. This was a SiVo hearing aid as a platform to present the tones. The earpiece of the SiVo (Oticon CP100) was coupled to a Bruel and Kjaer type 4157 ear simulator with type 4134 microphone, whose output was fed to a Rogers A75 amplifier and a pair of Connevans CE-8 earphones.

The SiVo was acoustically calibrated to be able to convert the tone presentation levels to relative HL values. The final frequency-shaping filter as derived from the POGO fitting algorithm was also stored on the SiVo-III.

The subject, who adjusted the output until the stimuli appeared comfortable, determined the final overall level. This level was set using a selection of stimuli from the "Flat" group and the same setting was used for all test conditions. At the beginning of each session, the level was re-checked using a selection of "Flat" tokens. Only a few tokens were required, so it was not believed that this initial stage gave an unfair advantage to the Flat processing condition.

Presentation of the stimuli was controlled via ASTEC software (Faulkner et al., 1997) and the PC video and sound card (Videologic Mediaspace) was used to play out the tokens.

The colour M-JPEG video images had 224 x 216 pixels, and were displayed on a 42.5 cm monitor with an overall screen resolution of 640 x 480 pixels at 25 frames per second. The video image occupied a screen area of approximately 21 by 20.5 cm. The talker's head (chin to top of head) had a vertical displayed height of 16 cm.

### *5.2.6 Task*

All test and practise lists for each processing condition contained all 45 tokens in a random order. Only one processing condition was tested in a session to avoid subjects getting confused between the different processing approaches. The order of presentation of processing conditions was randomized across subjects.

Once the frequency shaping had been loaded into the SiVo aid and the overall level had been adjusted, subjects were presented with two practice lists. The first list was an auditory-visual presentation to allow subjects to become familiar with the task, stimulus format and the use of the mouse. If necessary, a second auditory-visual list was presented to ensure that subjects were used to the task. Prior to testing, a final practise list was presented in auditory-alone format. This was done to allow the subjects to become used to attending to the auditory content of the stimuli. Feedback was given on a trial-by-trial basis.

Test runs were always presented in the following order:

Visual-alone, auditory-visual, auditory-alone, auditory-alone, auditory-visual, visual-alone.

The auditory-alone conditions were of most interest but the other conditions were useful for maintaining the confidence of the subjects by using a relatively easy task, and also to assess if the auditory signal had any effect on performance in the auditory-visual condition compared to the lipreading-alone condition.

The subjects were required to use the mouse to point to a button on the screen and to respond by labelling the fricative as either F, S, or SH. The written instructions to subjects were as follows:

*The sounds you will hear will be one of the following:*

| | | |
|---|---|---|
| *afa* | *asa* | *asha* |
| *ifi* | *isi* | *ishi* |
| *ufu* | *usu* | *ushu* |

*Concentrate on the consonant and ignore the vowel. Your task is to label which consonant you think you hear from a choice of three: F, S, SH.*

*The tests are carried out using a computer and you have to use a mouse to move the arrow on screen to the correct response and click the left button. The correct response will be shown once you have made your choice. There is no time limit. The next sound is only presented once you have responded.*

*Some of the conditions are very difficult. Please do not worry if you find them hard. It is a test of the processing strategy and not of you.*

Subjects were required always to respond even if the stimuli did not sound like any of the options. They were given feedback indicating the correct response on each trial and, if they requested, were told the final score at the end of the run.

## 5.3 Results

Figure 2 shows the results for all five subjects and the mean scores for the group. For the individual subjects, the lipreading score was the average of eight test runs (i.e. two from each of the four sessions). The auditory-visual and auditory-alone results are shown for each of the four processing conditions. Each column shows the average score for two test runs. The vertical bars within each column indicate the overall range of the results. For the group results, the columns indicate the mean scores across all the subjects, and the error bars show the standard error across subjects.

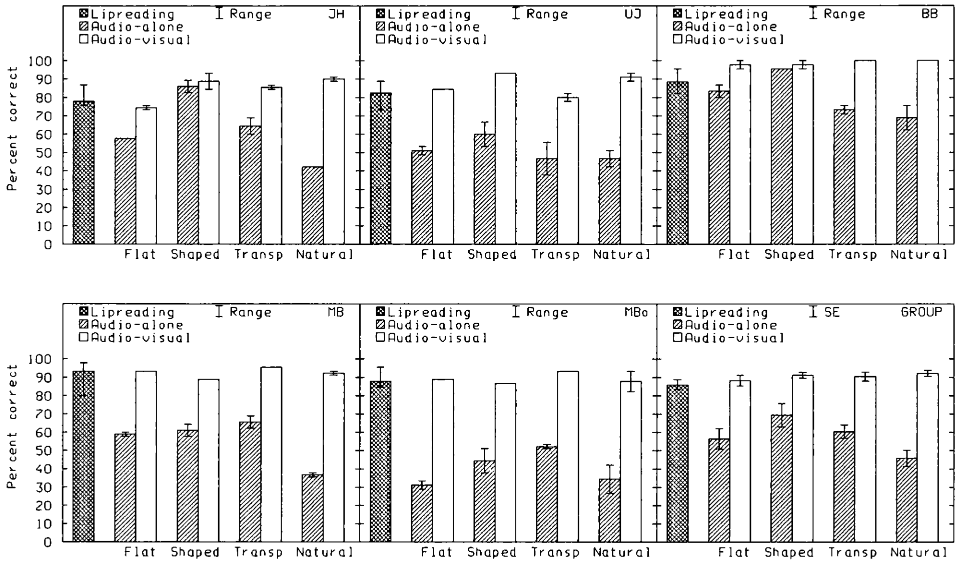



Figure 2. The results of individual subjects in all processing conditions. Five of the panels show the results for each subject and the bottom right panel shows the group results. The white bars show the auditory-visual performance and the hashed bars show the auditory-alone results. The black bar on the left-hand side of each panel shows the lipreading performance.

The natural condition was typically the one with the lowest scores. For subjects MB and MBo the scores were not above chance (33%). The scores for the shaped condition were the highest overall and subjects JH, UJ, and BB performed best with the shaped processing. Subject MBo had similar scores for shaped and transposed stimuli and subject MB scored similarly for the flat, shaped and transposed conditions. For the shaped condition, subjects JH and BB had high auditory-alone scores that were similar to the auditory-visual scores, indicating little dependence on visual cues. There was a large variation across subjects. Subjects BB and JH generally scored more highly than the other subjects in all conditions. This was probably due to them having a wider functional frequency range than the other subjects. Subject MBo never scored above approximately 50% for any of the auditory-alone conditions, whereas all the other subjects scored 60% or above in at least one auditory-alone condition. MBo had a greater hearing loss than the other subjects, with a severe hearing loss at low frequencies and a very limited frequency range. All of the other subjects had only moderate losses at 125 Hz. Subjects UJ and MB showed a severe loss at 250 Hz, and subjects BB and JH had severe losses by 500 Hz.

An arcsine transformation was carried out on all percent correct scores prior to statistical analysis. Two-way repeated measures analyses of variance (ANOVA's) were carried out on the combined results of the auditory-visual and visual-alone data to assess if there was any benefit of the auditory signal when combined with lipreading. A similar ANOVA was carried out on the combined data of the auditory-alone and auditory-visual conditions to assess if performance in the auditory-alone condition was improved with the addition of the visual cues.

The first analysis (auditory-visual vs. visual alone) with factors of processing condition and mode of presentation did not reveal any significant main effects or interactions. This means that subjects did not receive any significant extra information from the auditory signal over and above what they received from the visual information alone. This result was expected because the stimuli used were easily discriminable using visual cues alone. It is difficult to determine the effect the processing would have on auditory-visual speech perception because ceiling effects were observed for the visual alone performance.

The second analysis (auditory-visual vs. auditory-alone) with factors condition and mode of presentation revealed a significant effect of mode of presentation [$F(1,4) = 73.45$, $p = 0.001$] and a condition by mode interaction [$F(3,12) = 6.91$, $p = 0.006$]. Because of this interaction, one-way repeated measures ANOVA's were carried out on the auditory-alone and the auditory-visual data separately, with processing condition as a factor. The analysis of the auditory-alone data showed a significant effect of condition [$F(3,12) = 6.50$, $p = 0.007$]. The analysis of the auditory-visual data did not indicate a significant effect of condition ($F(3,12) = 1.10$, $p = .387$).

Two post-hoc tests were carried out for the auditory-alone analysis to determine if any of the conditions differed significantly. The first was the "Fisher's Least Significant Difference" test. It was determined that a difference of 0.267 (arcsine transformed values) was necessary for the difference to be significant at the 5% level. This gave two pairs with significant differences. These were the "flat" and "shaped" conditions, and the "shaped" and "natural" conditions. Shaped elicited the highest scores. Using the more conservative "Tukey's Honestly Significant Difference" test it was determined that a difference in means of 0.482 (in arcsine transformed values) was necessary for the difference to be significant at the 5% level. The only pairwise comparison that had a greater difference than this was that of "shaped" versus "natural" (difference of means = 0.537).

Finally, all of the subjects commented on the naturalness of the fricatives in the condition where they scored highest. This is an interesting response because it had been a long time since any of the subjects had heard fricatives naturally (between 25 and 50 years). One subject, JH, had been using a frequency transposing hearing aid (FreD, Frequency recoding device) on a daily basis. She was one of the better performers in the experiment, which may have been due to this exposure to transposed speech or due to having greater functional hearing than the other subjects. Certainly subjects JH and BB who were the subjects with the less severe losses preferred the "shaped" condition in terms of naturalness and this was also reflected in their performance. It is possible that naturalness in itself was a contributing factor to the performance of the subjects. In the conditions where the subjects perceived the stimuli to sound natural, they could have found the task easier to do because it was easier to provide the appropriate response categories to the correct stimuli. It doesn't necessarily mean that the subjects found the stimuli more discriminable in the shaped condition than the other conditions.

### 5.4 Discussion

In general, the subjects performed moderately well in this identification experiment. All subjects were able to use some of the auditory information in at least one of the processing conditions. This result is promising; especially considering the limited amount of exposure to the stimuli the subjects had received prior to testing. Also, it should be noted that subjects had to switch to a different processing algorithm for each new session of testing, which could have impaired performance. With regular use and auditory training it is possible that scores would have been much higher. The work of Rosen et al. (1999), described in chapter 2, looked at the effect of training on the perception of upward spectrally shifted speech in normal hearing listeners, as part of a study to simulate the speech processing received by cochlear implant users. They found that after receiving 3 hours of training using connected discourse tracking (De Filippo and Scott, 1978) the subjects showed a dramatic improvement in performance, from 1% to 30% with shifted BKB sentences. This shows that spectral distortions can pose major problems for speech perception in the first instance, but with training many of these difficulties can be overcome. This is encouraging for the current study, because subjects did not receive much training prior to performing the fricative perception task. It is possible that if subjects were given more exposure to the stimuli on a regular basis, or more training, then performance could have improved.

A feature of the project was to limit the number of stimuli to allow a range of processing conditions to be tested, and to give the subjects the best possible chance of performing at a high level in the test. The severe-to-profoundly hearing-impaired subjects tested here were reliant upon visual cues for speech perception in everyday situations and found testing auditory alone a very difficult task. For this reason only a small subset of those obstruents that contain high frequency energy were used and subjects were given feedback on a trial-by-trial basis. Many of the other obstruents (plosives, affricates and voiced fricatives) are typically shorter than the voiceless fricatives tested here and would have been more difficult to identify because of this. However in future work a smaller subset of processing conditions would be appropriate so a wider range of consonants could be tested. Also, there was no attempt here to address the issue of how well the subjects might do with additional vowel information, and whether they could process vowels presented naturally. It is possible that, with some processing of formant transitions, subjects would perform even better than currently found. A simplification of formant transition cues would facilitate the encoding of plosive place as well as fricative place. Stevens and Blumstein (1978) highlighted the importance of formant transitions for cueing place information in plosives, the transitions alone are sufficient to encode place whereas the burst alone can is not as and effective cue for giving place information. The addition of formant transition encoding should assist with differentiating the different manner of articulation classes, especially when the short duration of the voiceless excitation was not sufficient to allow subjects to extract a simplified spectral representation. This issue was not addressed in this experiment because the aim was solely to determine if information regarding the spectral differences of voiceless fricatives could be transmitted using noise bands alone. In future studies an open response set would be advisable to determine if the manner of articulation of the stimuli could be perceived. In the current design, as the aim was to determine if spectral differences could be signalled, the confusions with other manner classes was not evaluated. This would be an important consideration because many of these subjects have difficulty perceiving the difference between plosives and fricatives in natural speech (appendix 4).

The results of the statistical analyses showed that, for the auditory-alone processing mode, there was a significant effect of condition. The shaped processing condition gave significantly higher scores than the natural condition and than the flat condition. The shaped scores were the highest of all the scores, with the scores in the transposed condition next. However the difference between scores for the shaped and transposed conditions was not significant for the group. Neither was the score for the transposed condition significantly different from those for any of the other conditions. The flat processing condition gave the next lowest score, with the worst performance seen in the natural condition when stimuli were processed with frequency shaping according to the POGO gain rules.

It is difficult to know if subjects used both bandwidth and spectral peak location as cues for labelling the stimuli in the shaped processing condition, or whether they used just spectral peak location. However we do know that they were not basing responses purely on bandwidth, as this would have produced scores similar to those for the Flat processing condition. The listeners tested here also did not have good hearing at low frequencies. Thus, probably they were affected by impaired frequency selectivity. Subjects could have made use of the relative amplitude cues for signalling the sibilant/non-sibilant distinction but this would have held across all the processing conditions containing the auditory signal.

The statistical analyses indicated that there was no difference between visual alone and auditory-visual performance. The stimuli were relatively easy to discriminate based on visual cues alone, eliciting scores close to ceiling. This finding indicates that although the addition of the auditory signal to the visual signal did not improve performance, neither did it cause deterioration in performance, suggesting that the auditory signal did not cause difficulties in auditory-visual integration of the acoustic information to fricative place.

Overall, auditory-visual performance was significantly better than auditory-alone performance. Again, this is as expected, because the stimuli were easy to discriminate visually. The significant interaction of mode with processing condition reflects the significant effect of condition in the auditory-alone mode, and not the auditory-visual mode.

It is likely that the subjects tested here had dead regions at high frequencies due to the severity of the losses. Even if they did not have dead regions, conventional amplification could not give sufficient audibility for good speech perception. It is therefore necessary to find speech processing schemes that provide high-frequency speech information to these listeners, particularly when the low-frequency hearing is not sufficiently adequate to use cues from vowels to identify consonantal place information. This experiment showed that one potential approach for encoding fricatives would be to represent the original frication by a band of noise with the same relative bandwidth as in the original speech and shaped with one spectral peak at the same relative spectral location as in the original sound, but mapped down in frequency to within the listeners residual hearing range.

We anticipate that, in a final processing scheme, the degree of frequency mapping and dynamic range mapping required would be based upon the residual auditory abilities of the individual listener. Subjects with larger dynamic ranges could perhaps make use of bigger differences in relative level for coding features such as sibilance. Also, there could be more flexibility in the degree of frequency compression given to the individual subjects. The initial fitting could involve presenting a range of mapped stimuli to the listener and the choice of amount of compression could be based upon the individual's abilities to discriminate the stimuli.

# CHAPTER SIX

## 6 CONCLUSIONS

### *6.1* Introduction

The majority of hearing-impaired listeners have difficulty utilising the acoustic cues available in the speech signal to identify different phonemes. This could either be due to degradation in suprathreshold processing due to reduced frequency selectivity, loudness recruitment or the presence of dead regions, or that the cues are not sufficiently audible to the listener. The extent of the problem is dependent upon a combination of both the degree of hearing impairment and the extent of the degradation in suprathreshold processing. For severe-to-profoundly hearing-impaired listeners the extent of the reduction in suprathreshold processing and audibility is typically very large causing them great difficulties with speech perception.

To overcome the deficit in auditory processing severe-to-profoundly hearing impaired listeners rely heavily upon contextual information to assist with speech perception and are also particularly reliant upon lipreading to determine the place of articulation and give some manner cues.

There have been many attempts to lighten the perceptual load for severe-to-profoundly impaired listeners by exploring different speech processing approaches to deliver speech information to support lipreading, and also approaches that have aimed at providing re-coded auditory cues for speech information otherwise accessible only through lipreading.

One approach looked at reducing the complexity of speech to provide important features that could not be delivered by lipreading. Such a reduction in complexity has been shown in some instances to improve auditory-visual speech perception (Rosen et al., 1987; Faulkner et al., 1992; Wei, 1993; Faulkner et al., 1997). A pitch-tracking sinusoid was used to represent voice pitch and supplement lipreading. Wei (1993) used a neural network processing algorithm to extract the pitch. Faulkner et al. (1997) found that the most benefit from this approach was obtained for speech in noise, using noise-resistant pitch extraction analysis to extract voicing cues and fundamental frequency.

One difficulty with reducing the complexity of speech is that it is difficult to define the most appropriate cues to represent a speech sound without a full understanding of what the hearing-impaired listener can utilize. It is unlikely that the cues most appropriate for normal-hearing listeners will be the same as those cues that should be given to hearing-impaired listeners. It is also possible that within the hearing-impaired group some subjects will perform better with different cues to others. It is therefore necessary that the cues used be carefully selected with the opportunity to provide different cues to different listeners dependant upon their processing capabilities.

Another approach to assisting speech perception for hearing-impaired listeners is to shift the high-frequency speech energy into the low-frequency region to provide the speech cues that are inaudible to the hearing-impaired listener. This has been extensively investigated (e.g. Johannson, 1961; Velmans, 1973; Rosenhouse, 1990; McDermott and Dean 2000) and there has been very little evidence showing success. It is difficult to determine why the approach has been unsuccessful because a wide range of processing strategies has been attempted. Some schemes increase the complexity of the low frequency information, which may overload the impaired auditory system; some approaches affect the pitch of the speech so may sound very unnatural; some approaches map the information in a way that might distort the transposed information. It might be that information shifted down in frequency cannot be integrated with the low-frequency information and as a result the signal forms two streams of information (however this has never been reported).

Frequency transposition has the potential of representing the voiceless cues for fricative perception that are present in the high frequencies. One of the difficulties with voiceless excitation is that it is usually inaudible for most hearing-impaired listeners. This is because the relative level of the excitation is weak compared to that of vowels and also the frequency content of the voiceless excitation is typically in the high frequencies outside the frequency-range of hearing of many hearing-impaired listeners.

There is evidence that if the voiceless excitation can be made audible then this is sufficient for identifying these speech sounds (Hogan and Turner 1998). Unfortunately it is not possible to just amplify voiceless excitation to audible levels for listeners with severe-to-profound losses because the high frequency regions of their hearing are often dead (Moore et al. 2001). Therefore, for these listeners the most appropriate approach is to shift the frequency components down to the region of functional hearing in the listener.

It may be the case that frequency compression or transposition alone is not sufficient to improve speech perception and that the complexity of the signal needs to be reduced to allow the hearing-impaired listener to utilise the cues. Therefore an approach combining speech pattern extraction and frequency compression may be most suitable.

The primary goal of the work in this thesis was to determine whether the spectral cues for place of information inherent in the voiceless excitation of fricatives could be encoded to allow severe-to-profoundly hearing-impaired listeners to perceive place information without relying on lipreading. A selection of experiments was conducted addressing different aspects of this question.

## 6.2 Overview of work in this thesis

The initial experimental chapter looked at whether fricatives could be simplified for normally hearing listeners and still retain important information about the place of articulation for the fricatives. It was important that the simplifications of the fricatives remained as true to the original token as possible even though simplified. To address this the pattern of results were interpreted in relation to the normal perception of natural fricatives: if the simplifications caused the listener to use completely different labelling parameters to those used for normal fricatives then this would have been unacceptable. Also it was important that the subjects all followed similar patterns for labelling the simplified tokens to ensure that individual subjects were not just selecting specific arbitrary labels for the different tokens.

In the experiment the normal-hearing subjects listened to band-pass filtered noises. The noises varied in bandwidth, relative level and centre frequency. The subjects had to label the noises with one of five voiceless fricative labels. The pattern of responses followed the general pattern of responses described by other researchers (Harris, 1958; Heinz and Stevens, 1961) and those that would be predicted from spectrographic analysis (Strevens, 1960). The subjects were able to label the sounds as one or the other sibilants /s, ʃ/ based on the static noise characteristics, but they did not differentially label the non-sibilants (f, θ) with these static cues. This outcome supports the idea that formant transitions, which were not present in these stimuli, play a big role in cueing non-sibilant place. The choice of the sibilant/non-sibilant categories was based on the relative level and bandwidth of the noises. The narrow bandwidth sounds with high relative level were generally given sibilant labels and the broader noise bands with low relative level were labelled as non-sibilant. /f/ was typically given as a response for high-frequency tokens with the latter characteristics. The centre frequency (CF) of the noise band was the primary determinant of the chosen place of articulation for the sibilants.

Low CF's were labelled as /ʃ/ and high frequencies as /s/. The variability between subjects in their choice of labels was relatively small (the group mean standard deviation was 1.92, collapsed across all the conditions). The results showed that very simple noises could be used to carry some of the critical information required for labelling voiceless fricatives while still retaining some of the natural qualities used for labelling natural fricatives by normal hearing listeners.

The second and third experimental chapters described experiments to determine the number of noise bands that could be discriminated within the residual hearing range at durations appropriate for running speech and also to determine if bandwidth was a cue that could be used by measuring psychometric functions for bandwidth discrimination.

All subjects tested could discriminate between at least two 250 Hz-wide noise bands at durations of 80 ms or less when the noises were spectrally adjacent (i.e. the upper-frequency edge of one was the same as the lower-frequency edge of the other). Furthermore, subjects could discriminate the bandwidth of noise stimuli reasonably well when the noises had the same centre frequency: thresholds typically corresponded to bandwidth ratios between 1.3:1 and 2:1. This indicates that it may be possible to encode fricatives as noise bands for these listeners, because the perceptual abilities required to discriminate such encoded sounds are present.

Although the psychophysical work did not explore the intensity discrimination skills of the listeners it was expected that the majority of the listeners would be able to discriminate at least two intensity steps within their dynamic range. It was hoped that this discrimination ability could be exploited for representation of the sibilant/non-sibilant categories within the limited dynamic hearing range of the listener. The sibilants could be represented at the upper limit of the residual intensity range and the non-sibilants towards the lower limit ensuring that the non-sibilants were audible and the sibilants were not uncomfortable.

In the final experimental chapter a study was conducted using a range of simplification approaches to encode fricatives, and a group of severe-to-profoundly hearing-impaired listeners were asked to label the encoded stimuli as fricatives.

The subjects listened to presentations of VCV syllables where the consonant was one of /f/, /s/, or /ʃ/. The vowel environment was either /i/, /a/, or /u/. The vowels were represented by a sinusoid tracking the fundamental frequency of the vowel and the consonants were processed using one of four strategies differing in complexity: The flat processing scheme was intended to be the simplest approach with just a simple band-pass noise representing the sound. The shaped stimuli were slightly more complex with the spectra shaped with one spectral peak. A frequency-compressed condition very similar to simple transposition had many spectral peaks. The fourth control condition used unprocessed fricative noise.

The results showed that, for the auditory-alone perception, there was a significant effect of condition. This was interesting because the aim of the experiment was to assess if speech processing could give sufficient auditory information to identify voiceless fricatives in the absence of visual cues. The shaped processing condition gave the highest scores, with the scores in the transposed condition next. The flat processing condition gave the next lowest score, with the worst performance seen in the unprocessed condition.

The results of this experiment were very encouraging because these subjects are typically highly reliant on lipreading to be able to perceive place information in everyday life. Three of the subjects had been tested on a 24 consonant VCV test (results shown in appendix?) and on average only scored 21.5%. When just observing the scores for just the /f/, /s/ and /ʃ/ responses the group results were 13 out of 48 correct a percentage of 27%. For the shaped condition in the final experiment the mean score was approximately 70%, which is a large improvement.

Some caution must be shown when interpreting the results because the subjects in this experiment were only given three response options whereas when tested on the 24 VCV task there were 24 response options available. The results with 24 VCV's showed that two of the subjects made quite a few manner errors, this was not a possibility in the simplified fricative identification experiment because only fricative response categories were available. Further experimentation is required to determine if by encoding the obstruents into the low frequencies it will also give improvements in manner distinctions. This could be a possibility because with extra cues available it could assist with identifying manner as well as place of articulation.

The outcomes from this experiment indicate that there could be some potential for encoding fricatives by representing the original frication by a band of noise with the same relative bandwidth as in the original speech and shaped with one spectral peak at the same relative spectral location as in the original sound but mapped down in frequency to within the listeners residual hearing range. Further work will show if this encoding is also appropriate for other obstruent classes.

This thesis focussed on one small component of the speech signal, the voiceless fricatives. The aim was to see if severe-to-profoundly hearing-impaired listeners retained hearing abilities that were sufficient for identifying encoded voiceless fricatives represented by low-frequency bands of noise. Although the work focussed on a small subset of speech sounds it was intended that encoding strategies would be derived that could potentially be used more broadly to encode spectral features of all voiceless speech sounds. The results of the experiments carried out are encouraging and should prompt further work in the area.

## 6.3 Further work

It would be interesting to develop the work to look at the perception of other phoneme groups, namely plosives and affricates and also to assess if this processing assists when listening to sentences rather than isolated VCV's. It would be important to assess the effect of the processing on general sound quality, e.g., music perception, environmental sounds and voice quality with a range of different voices.

It would be interesting to test the shaped, transposed and flat processing in conjunction with natural vowels or perhaps simplified representations where just one formant is presented and see how the perception of the processed phonemes is affected.

The ultimate aim of future work should be to provide a processing approach that runs in real-time, tracking the dynamic changes in natural speech and encoding all voiceless excitation. A systematic approach to the encoding of fricatives by relating the encoded versions to the bandwidth, intensity and main peak frequency of the original frication would seem most appropriate at this stage.

The final processing scheme should allow flexibility to adjust the degree of frequency mapping and dynamic range mapping required to match the residual auditory abilities of the individual listener. Subjects with larger dynamic ranges could perhaps make use of bigger differences in relative level for coding features such as sibilance. Also, there could be more flexibility in the degree of frequency compression given to the individual subjects. The initial fitting could involve presenting a range of mapped stimuli to the listener and the choice of amount of compression could be based upon the individual's abilities to discriminate the stimuli.

It must not be forgotten that many severely and profoundly hearing-impaired listeners will be eligible to receive a cochlear implant, in fact three of the subjects participating in this work have since received implants and are performing well with their devices. However this does not exclude further work within this area because there are many listeners who for medical, financial or personal reasons are not suitable for cochlear implantation, or have to wait a lengthy period of time prior to implantation. It is therefore necessary to fully explore the range of processing options available for improving speech perception acoustically for these listeners.

## 6.4 Conclusion

Work has been carried out to show that there is some potential for the simplification and transposition of voiceless excitation in speech to improve the identification of voiceless fricatives for severe-to-profoundly hearing-impaired listeners. Simple noise bands varying in bandwidth, level and centre frequency can transmit the main cues for fricative perception. For hearing-impaired listeners these noise bands can be labelled as fricatives when they are placed within the residual hearing range of the listener. Further work is required to define an encoding strategy adapted for all voiceless excitation and to create real-time processing to work with natural speech.

# 7 References


Bernstein, L. E., Demorest, M. E. and Tucker, P. E. (2000) Speech perception without hearing. Percept Psychophys **62(2)**:233-52

Bilger, R. and M. D. Wang (1976). "Consonant confusions in patients with sensorineural hearing loss." Journal of Speech and Hearing Research **19**: 718-748.

Borg, E., B., Canlon, B. and Engstrom, B. (1995). "Noise-induced hearing loss - Literature review and experiments in rabbits." Scandinavian Audiology **24, Suppl. 40**: 1-147.

Borden G., Harris K. (1984). "Speech Science Primer," Williams and Wilkins, Baltimore/London.

Braida, L.D., Durlach, N.I., Lippmann, R.P, Hicks, B.L., Rabinowitz, W.M. and Reed, C.M. (1979). "Hearing aids - a review of past research on linear amplification, amplitude compression, and frequency lowering." ASHA Monogr. 1979 Apr;(19):1-114.

Davis, A. (1995). Hearing in Adults. London, Whurr.

De Filippo, C. L. and Scott, B. L. (1978). "A Method for training and evaluating reception of ongoing speech. J. Acoust. Soc. Am. **63**: 1186-1192.

Dorman, M. F. and P. C. Loizou (1998). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using stimulations of processors with two to nine channels." Ear and Hearing **19**(2): 162-165.

Drullman, R. and Smoorenburg, G. F. (1997). "Audio-visual perception of compressed speech by profoundly hearing-impaired subjects". Audiology **36:** 165-177

Fabry, D. A. and D. J. van Tasell (1986). "Masked and filtered simulation of hearing loss: Effects on consonant recognition." Journal of Speech and Hearing Research **29**: 170-178.

Farrar, C.L., Reed, C.M., Ito, Y., Durlach, N.I., Delhorne, L.A., Zurek, P.M., Braida, L. D. (1987). "Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises". J. Acoust. Soc. Am. **81**: 1085-1092.

Fastl, H. (1980). "Pitch strength and masking patterns of low-pass noise". *In Psychophysical, psychological and behavioural studies in hearing.* Edited by van der Brink, G and Bilsen, F.A. 334-340.

Fastl (1971). "Über Tonhöhenempfindungen bei Rauschen." Acustica **25**: 350 -354.

Faulkner, A and Wei, J. (1997). "Performance of a real-time multi-layer perceptron algorithm for the extraction of the voicing and voice fundamental frequency information." Speech, Hearing and Language, Work in progress, Department of Phonetics and Linguistics, University College London, Vol. 10 p 41 - 60.

Faulkner, A., Fourcin, A. J. and Moore, B. C. J. (1990). "Psychoacoustic aspects of speech pattern coding for the deaf." Acta Otolaryngol. Suppl. **469**: 172-180.

Faulkner, A., Rosen, S. and Moore, B. C. J. (1990). "Residual frequency selectivity in the profoundly hearing-impaired listener." Brit. J. Audiol. **24**: 381-392.

Faulkner, A. and Rosen, S. (1999). "Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception." J. Acoust. Soc. Am. **106**(4): 2063-2073.

Faulkner, A., van Son, N. and Beijk, C. (1997). "The TIDE project OSCAR: Scand. Audiol. 26 **Suppl. 47**: 38 - 44.

Finney, D. J. (1971). "Probit Analysis" Cambridge University Press. Edition 3.

Flynn, M. C., Dowell, R. C. and Clark, G. M. (1998). "Aided speech recognition abilities of adults with a severe or severe-to-profound hearing loss." Journal of Speech, Language and Hearing Research **41**: 285-299.

Fourcin, A. J. and Abberton, E. (1977). "Laryngograph studies of vocal-fold vibration." Phonetica **34**(4): 313-315.

Fowler, E. P. (1936). "A method for the early detection of otosclerosis." Archives of Otolaryngology **24**: 731-741.

Fu, Q. J., Shannon, R. V. and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing." J. Acoust. Soc. Am. **104**(6): 3586-3596.

Glasberg, B. R. and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments." J. Acoust. Soc. Am. **79**: 1020-1033.

Glasberg, B. R., Moore, B. C. J. and Bacon, S. P. (1987). "Gap detection and masking in hearing-impaired and normal-hearing subjects." J. Acoust. Soc. Am. **81**: 1546-1556.

Grant, K. W., Walden, B. E. and Seitz, P. F. (1998). "Auditory-visual speech recognition by hearing-impaired subjects : Consonant recognition, sentence recognition, and auditory-visual integration." J. Acoust. Soc. Am. **103**(5): 2677-2689.

Harris, J. D. (1952). "Pitch Discrimination". J. Acoust. Soc. Am. **24**: 750-755.

Harris, K. S. (1958). "Cues for the Discrimination of American English Fricatives in Spoken Syllables". Lang. Speech **1**: 1 - 7

Heinz, J. M. and Stevens, K. N. (1961). "On the properties of Voiceless Fricative Consonants". J. Acoust. Soc. Am. **33(5)**: 589-597.

Hogan, C. A. and Turner, C. W. (1998). "High-frequency audibility: Benefits for hearing-impaired listeners." J. Acoust. Soc. Am. **104**(1): 432-441.

Houtgast, T. (1972). "Psychophysical evidence for lateral inhibition in hearing." J. Acoust. Soc. Am. **51**: 1885-1894.

Howell, P. and Rosen, S. (1983). "Production and Perception of rise time in voiceless affricate/fricative distinction". J. Acoust. Soc. Am. **73**: 976-984

Johansson, B. (1966). "The use of the transposer for the management of the deaf child." International Audiology **5**: 362-372.

Kamm, C. A., Dirks, D. D. and Bell, T. S. (1985). "Speech recognition and the Articulation Index for normal and hearing-impaired listeners." J. Acoust. Soc. Am. 77: 281-288.

Klatt, (1980). "Software for a Cascade/Parrallel Formant Synthesizer". J. Acoust. Soc. Am. **67**: 971-995

Lawrence and Byers (1969). "Identification of voiceless fricatives by high frequency hearing impaired listeners". Journal of Speech and Hearing Research **12**: 426-434.

Levitt, H. (1971). "Transformed Up-Down Methods in Psychoacoustics". J. Acoust. Soc. Am. **49**: 467-477

McCandless and Lyregaard (1983). "Prescription of Gain/Output (POGO) for Hearing Aids." Hearing Instruments **34:** 16-21

McDermott, H. J. and Dean, M. J. (2000). "Speech perception with steeply sloping hearing loss: Effects of frequency transposition." Brit. J. Audiol. **34(6)**: 353-361.

McDermott, H. J., Dorkos, V. P., Dean, M. R. and Ching, T.Y. (1999). "Improvements in speech perception with use of the AVR TranSonic frequency-transposing hearing aid." Journal of Speech, Language and Hearing Research **42**: 1323-1335.

McGurk, H. and McDonald J. (1976). "Hearing lips and seeing voices." Nature **264**: 746-748.

Miller, G. A. and Nicely P. E. (1955). "An analysis of perceptual confusions among some English consonants." J. Acoust. Soc. Am. **27**: 338-352.

Moore, B. C. J. (1998). Cochlear Hearing Loss. London, Whurr.

Moore, B. C. J. and Glasberg, B. R, (1998). "The use of a loudness model for hearing aid fitting. I. Linear hearing aids. Br. J. Audiol; **32**: 289-311

Moore, B. C. J. and Glasberg, B. R. (1986). "The relationship between frequency selectivity and frequency discrimination for subjects with unilateral and bilateral cochlear impairments," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson. (Plenum, New York).

Moore, B. C. J. and Glasberg B.R. (1988). Practical and theoretical considerations in designing and implementing automatic gain control (AGC) in hearing aids. Clinical Audiology '87. A. Quaranta. Bari, Italy, Laterza.

Moore, B. C. J., Huss, M., Vickers D.A., and Baer, T. (2001). Psychoacoustics of dead regions. Physiological and Psychophysical Bases of Auditory Function. A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs and R. Schoonhoven. Maastricht, Shaker: (in press).

Moore, B.C.J., Huss, M., Vickers, D. A., Glasberg, B. R. and Alcántara, J.I. (2000). *A test for the diagnosis of dead regions in the cochlea.* British Journal of Audiology, **34**: p. 205-224.

Moore, B. C. J., Oldfield, S. R. and Dooley, G. (1989). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz." J. Acoust. Soc. Am. **85**: 820-836.

Moore, B. C. J., Vickers, D. A., Glasberg, B. R. and Baer, T. (1997). "Comparison of real and simulated hearing impairment in subjects with unilateral and bilateral cochlear hearing loss". Brit. J. Audiol. **31**: 227-245.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (1996). "Effect of loudness recruitment on the perception of amplitude modulation". J. Acoust. Soc. Am. **100 (1)**: 481-489.

Moore, B. C. J., and Glasberg, B. R. (1997). "A model of loudness perception applied to cochlear hearing loss," Aud. Neuro. 3, 289 - 311.

Murray, N. and D. Byrne (1986). "Performance of hearing-impaired and normal hearing listeners with various high-frequency cut-offs in hearing aids." Australian Journal of Audiology **8**: 21-28.

Osberger, M. J., Zimmerman-Phillips, S., and Koch, D. B. (2002). "Cochlear Implant candidacy and performance trends in children". Annals of Otology Rhinology and Laryngology, **111**: 62-65.

Patterson, R. D. and Nimmo-Smith, I. (1980). "Off-frequency listening and the auditory filter assymmetry." J. Acoust. Soc. Am. **67**: 229-245.

Pavlovic, C. V. (1984). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment." J. Acoust. Soc. Am. **75**: 1253-1258.

Pick, G. F., Evans, E. F., Wilson, J. P. (1977). "Frequency resolution in patients with hearing loss of cochlear origin," in *Psychophysics and physiology of hearing,* edited by E. F. Evans and J. P. Wilson (Academic, London).

Pickett J. M., Daly R. L., and Brand S. L. (1965). "Discrimination of spectral cutoff frequency in residual hearing and in normal hearing". J. Acoust. Soc. Am. **38:** 923.

Plomp, R. (1964). "The ear as a frequency analyzer." J. Acoust. Soc. Am. **36**: 1628-1636.

Posen, M. P., Reed, C. M. and Braida, L. D. (1993). "Intelligibility of frequency-lowered speech produced by a channel vocoder." Journal of Rehabiliation Research and Development **30**(1): 26-38.

Reed, C. M., Power, M. H., Durlach, N. I., Braida, L. D., Foss, K. K., Reid, J. A. and Dubois, S. R. (1991). "Development and testing of artificial low-frequency speech codes." Journal of Rehabiliation Research and Development **28**(3): 67-82.

Rosen, S. (1992). "Temporal information in speech: acoustic, auditory and linguistic aspects". Philosophical Transactions of the Royal Society of London. Series B-336, **1278**: 367-373.

Rosen, S., Faulkner, A. and Smith, D. A. J. (1990). "The Psychoacoustics of Profound Hearing Impairment." Acta Otolaryngol (Stockh) **Suppl. 469**: 16-22.

Rosen, S., Faulkner, A. and Wilkinson, L. (1999). "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants". J. Acoust. Soc. Am. **106(6)**: 3629-36.

Rosen, S. and Smith, D. (1987). "Temporally-based auditory sensations in the profoundly hearing-impaired listener," in *Basic Issues In Hearing* edited by H. Duifhuis, J. W. Horst, and H. P. Wit. (Academic Press, London).

Rosen, S. and Stock, D. (1992). "Auditory filter bandwidths as a function of level at low frequencies (125 Hz-1 kHz)." J. Acoust. Soc. Am. **92**: 773-781.

Rosen, S., Walliker, J. R., Fourcin, A. J., and Ball, V. (1987). "A microprocessor-based acoustic hearing aid for the profoundly impaired listener," J. Rehabil. Res. Devel. **24**: 239 – 260.

Rosenhouse, J. (1990). A new transposition device for the deaf. The Hearing Journal, **43:** 20-25.

Shailer, M. J. and B. C. J. Moore (1983). "Gap detection as a function of frequency, bandwidth and level." J. Acoust. Soc. Am. **74**: 467-473.

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

Sharma, A., Dorman, M., Spahr, A., and Todd, N.W. (2002). "Early cochlear implantation in children allows normal development of the central auditory pathways". Annals of Otology Rhinology and Laryngology, **11:**, 38-41.

Skinner, M. W. and J. D. Miller (1983). "Amplification bandwidth and intelligibility of speech in quiet and noise for listeners with sensorineural hearing loss." Audiology **22**: 253-279.

Skinner, M. W., Holden, L. K., and Binzer, S. M. (1994). "Aural rehabilitation for individuals with severe and profound hearing impairment: Hearing Aids, Cochlear Implants, Counselling and Training." In *Strategies for selecting and verifying hearing aid fittings.* Thieme Medical Publishers. 249-266

Souza, P. E. and Turner, C. W. (1999). "Quantifying the contribution of audibility to recognition of compression-amplified speech." Ear and Hearing **20**(1): 12-20.

Stevens, K. N. and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants". J. Acoust. Soc. Am. **64(5)**: 1358-1368.

Strevens (1960). "Spectra of Fricative Noise in Human Speech". Lang. Speech **3**: 32 - 49

Summerfield, Q. (1987). "Some preliminaries to a comprehensive account of audio-visual speech perception." In Hearing by Eye: The psychology of Lipreading. (Eds Dodd, B. and Campbell, R.), Erlbaum, New Jersey.

van Tasell, D. J., Fabry, D. A., and Thibodeau, L. M. (1987). “Vowel identification and vowel masking patterns of hearing-impaired subjects,” J. Acoust. Soc. Am. **81 (5):** 1586 - 1597.

Turner, C. W. (1999). "The limits of high-frequency amplification." Hearing Journal **52**(2): 10-14.

Turner, C. W. and Cummings, K. J. (1999). "Speech audibility for listeners with high-frequency hearing loss." American Journal of Audiology **8**: 47-56.

Velmans, M. (1974). "The design of speech recoding devices for the deaf". Brit. J. Audiol. **8:** 1-15.

Versfeld, N. J. (1997). "Discrimination of changes in spectral shape of noise bands". J. Acoust. Soc. Am. **102** (5): 1586 - 1597.

Vickers, D.A., Moore. B. C. J., and Baer, T. (2001) *Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies.* Journal of the Acoustical Society of America. **110**(2): p. 1164-1175

Walden, B. E., Schwartz, D. M., Montgomery, A. A. and Prosek, R. A. (1981). "A comparison of the effects of hearing impairment and acoustic filtering on consonant recognition." Journal of Speech and Hearing Research **24**: 32-43.

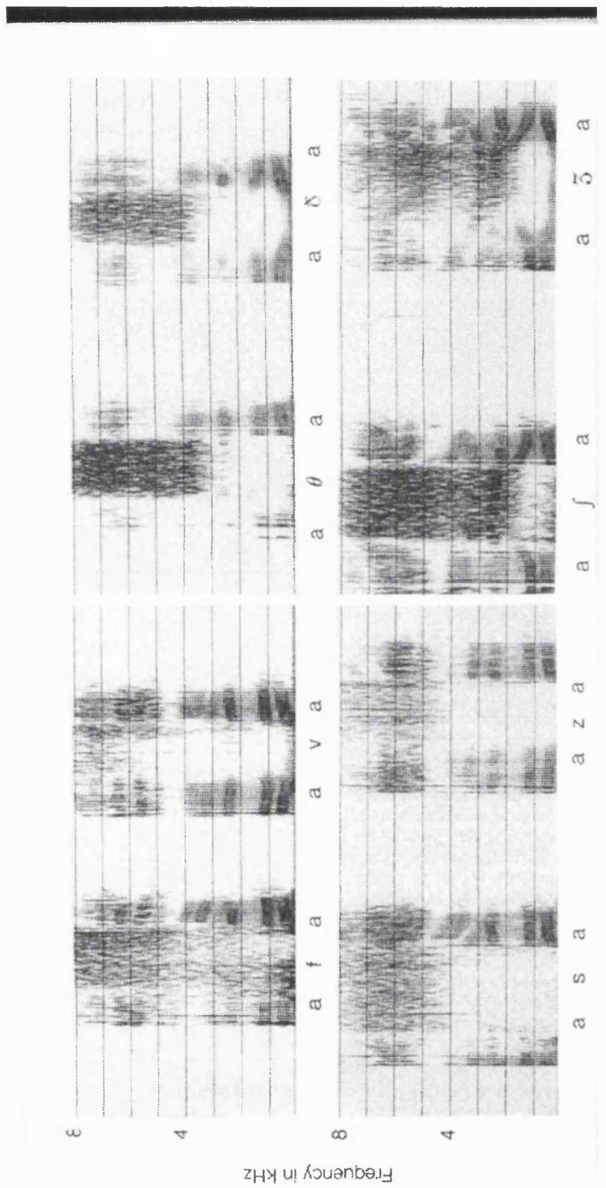
Walliker, J. R. and Howard, I. S. (1990). "The implementation of a real time speech fundamental period algorithm using multi-layer perceptrons." Speech Communication **9**: 63-71.

Wei, J. (1993). "A speech-Pattern Processing Method for Chinese Listeners with Profound hearing Loss". PhD submitted to the Department of Phonetics and Linguistics, University College London

Warren, R. M. and Bashford, J. A. (1999). "Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband." J. Acoust. Soc. Am. **106**(5): L47-L52.

Zeng, F.-G. and Turner, C. W. (1990). "Recognition of voiceless fricatives by normal and hearing-impaired subjects." Journal of Speech and Hearing Research **33**: 440-449.

# 8 Appendix 1. Spectrograms of natural fricatives. Courtesy of Borden and Harris (1984).



# 9 Appendix 2. Analysis of stimuli used in Chapter 2 "The labelling of bandpass filtered noise bands as fricatives by normal hearing listeners".

The table below shows the results of measurements made on the stimuli used in chapter two. Two groups of the stimuli were used to make the calculations these were the "0dB" and the "-6dB" relative level groups. The purpose of the analysis was to assess if the peak levels of the spectra of the noise bands changed with centre frequency (CF) and bandwidth.

The analyses were carried out using the Cool Edit 2000 programme from Syntrillium Corporation. For all the stimuli listed in the table the noise band was selected and the Average RMS was calculated. In addition to this a long-term average spectrum was produced of the same section of the band and the level of the peak was measured. The same frequency point was used for the equivalent stimulus in both the 0dB and –6dB sets. The RMS was computed over the entire frequency range and the spectral peak level was taken from the frequency range at the peak. The 0dB point for the Spectral Peak level is the highest point on the analysis range. The FFT size was 2048 and the sampling rate was 20,000 Hz.

The table shows that the RMS level remains similar across the stimuli and the peak level increases as bandwidth decreases and it also decreases with increasing CF. This is due to the radiation characteristic of the Klatt synthesiser. The 6 dB level difference can be clearly seen when looking at the two stimulus sets.

| Bandwidth | CF | Rel Level (dB) | Spectral Peak Level | Mean RMS | Bandwidth | CF | Rel Level (dB) | Spectral Peak Level | Mean RMS |
|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 1 | 0 | -42.19 | -26.43 | 0.5 | 1 | -6 | -48.59 | -32.53 |
| 0.5 | 2 | 0 | -42.14 | -26.34 | 0.5 | 2 | -6 | -48.26 | -32.43 |
| 0.5 | 3 | 0 | -43.45 | -27.07 | 0.5 | 3 | -6 | -49.58 | -33.16 |
| 0.5 | 4 | 0 | -45.32 | -26.59 | 0.5 | 4 | -6 | -50.32 | -32.62 |
| 0.5 | 5 | 0 | -45.14 | -26.31 | 0.5 | 5 | -6 | -51.43 | -32.68 |
| 0.5 | 6 | 0 | -43.94 | -26.35 | 0.5 | 6 | -6 | -51.34 | -32.40 |
| 0.5 | 7 | 0 | -45.27 | -26.35 | 0.5 | 7 | -6 | -51.67 | -32.44 |
| 1.0 | 1 | 0 | -43.80 | -26.62 | 1.0 | 1 | -6 | -50.73 | -32.72 |
| 1.0 | 2 | 0 | -44.94 | -26.61 | 1.0 | 2 | -6 | -51.09 | -32.70 |
| 1.0 | 3 | 0 | -46.22 | -27.19 | 1.0 | 3 | -6 | -52.23 | -33.29 |
| 1.0 | 4 | 0 | -47.74 | -27.03 | 1.0 | 4 | -6 | -52.92 | -32.83 |
| 1.0 | 5 | 0 | -47.93 | -26.50 | 1.0 | 5 | -6 | -53.93 | -33.13 |
| 1.0 | 6 | 0 | -47.70 | -26.50 | 1.0 | 6 | -6 | -53.54 | -32.56 |
| 1.5 | 1 | 0 | -44.81 | -26.42 | 1.5 | 1 | -6 | -50.88 | -32.59 |
| 1.5 | 2 | 0 | -45.42 | -26.24 | 1.5 | 2 | -6 | -51.48 | -32.33 |
| 1.5 | 3 | 0 | -47.11 | -26.83 | 1.5 | 3 | -6 | -53.22 | -32.94 |
| 1.5 | 4 | 0 | -47.49 | -26.49 | 1.5 | 4 | -6 | -53.57 | -32.58 |
| 1.5 | 5 | 0 | -48.03 | -26.13 | 1.5 | 5 | -6 | -53.83 | -32.21 |
| 2.0 | 1 | 0 | -44.39 | -26.50 | 2.0 | 1 | -6 | -51.17 | -32.60 |
| 2.0 | 2 | 0 | -45.42 | -26.31 | 2.0 | 2 | -6 | -52.48 | -32.40 |
| 2.0 | 3 | 0 | -47.75 | -26.92 | 2.0 | 3 | -6 | -53.86 | -33.02 |
| 2.0 | 4 | 0 | -48.12 | -26.69 | 2.0 | 4 | -6 | -54.54 | -32.79 |

**Table A Analysis of stimuli used in Chapter 2**

## 10 Appendix 3. Mean Standard Deviations separated in terms of response category and bandwidth

To calculate the mean standard deviations the scores out of ten for each of the 176 stimuli were determined. The average scores were then calculated for each stimulus mean and also the standard deviation. This was conducted for each response category (ʃ, s, f, θ). These scores have been reduced to show the mean standard deviation as a function of the bandwidth categories, i.e. the means that were produced for each stimulus have been averaged for the bandwidth group.

The average standard deviations are relatively small for the group. In general the subjects showed similar patterns of responses.

| | Bandwidth | | | |
|---|---|---|---|---|
| | **0.5** | **1.0** | **1.5** | **2.0** |
| **ʃ** | 1.68 | 1.75 | 2.04 | 2.20 |
| **s** | 2.07 | 2.16 | 2.29 | 2.24 |
| **f** | 1.56 | 1.65 | 1.73 | 1.89 |
| **θ** | 1.70 | 1.65 | 1.98 | 2.06 |

# 11 Appendix 4. Results from consonant identification tests for three of the five subjects tested in chapter five listening to natural speech using a conventional hearing aid auditory alone.

## 11.1 Aim

The data presented in the following table allows an assessment of the hearing-impaired subjects abilities in the identification of place of articulation in natural fricatives.

## 11.2 Method and Subjects

Data was available from three of the subjects, JH being one of the better performers and UJ and MB being typical of the poorer performing subjects.

The subjects were all tested using an optimally fitted conventional hearing aid without the use of lipreading. Stimuli were intervocalic consonants produced by a female British RP speaker. The stimuli were presented in a sound treated room, played from a U-Matic video recorder, through a Yamaha P2050 amplifier and presented free-field from a Quad PRO-63 electrostatic loudspeaker.

## 11.3 Results

Table I shows the results for subjects JH, MB and UJ in a 24 consonant intervocalic consonant test.

**Table I. Confusion Matrices for subjects JH, MB and UJ for stimuli /f/, /s/ and /ʃ/**
The listening environment and the well-pronounced stimuli provided optimal opportunity for consonant identification. It is most likely that in everyday life that these subjects would have performed more poorly than indicated in these tests.

**Subject JH –** had the best hearing of the group. She made no manner errors for the phonemes /f, s, ʃ/. However she made many place errors and even some voicing errors when confusing /s/ for /z/. Her overall percent correct identification score for consonant identification was still very low, most of this deficit being due to place errors (mean value was 39.6%), her manner perception score was relatively high (mean value of 70.8%).

**Subject MB –** was able to distinguish obstruents from other manner classes, but made many confusions between fricatives and plosives. She used /f/ as one of her preferred responses and made many place confusions. She had the lowest score of the group for overall percent correct for consonant recognition (13.55%) the poor identification was due to both manner errors (only 39.6% correct) and place errors (35.4% correct).

**Subject UJ –** UJ was able to identify the obstruent group but made many fricative/plosive confusions. She regularly used the /f/ response and made many place errors. She never used the /ʃ/ response category. Her overall percent correct score was 21.85% her difficulties were caused by both errors in manner (only 47.9% correct) and place perception (39.6% correct).

## 11.4 Conclusion

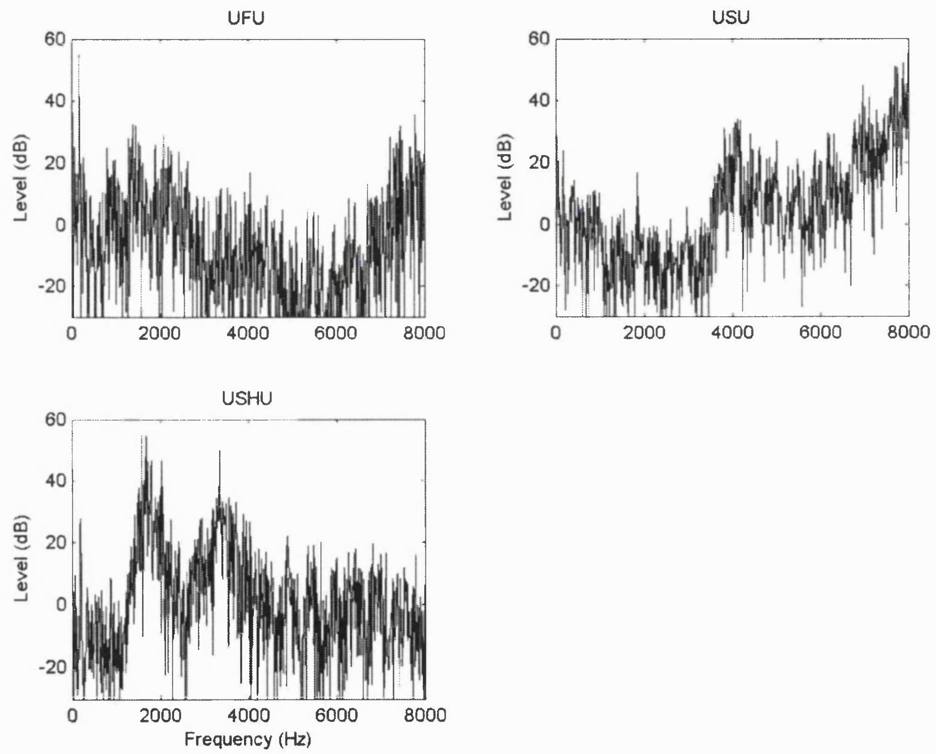
The results showed that the severe-to-profound hearing-impaired listeners used in the experiment in chapter five were unable to identify fricative place and often also made errors with plosive classes as well. To overcome these difficulties severe-to-profoundly hearing-impaired listeners are heavily reliant upon lipreading for speech perception.

# 12 Appendix 5. The spectra of the stimuli used in Chapter 5 "The labelling of noise bands as fricatives within a speech-like environment by hearing-impaired listeners

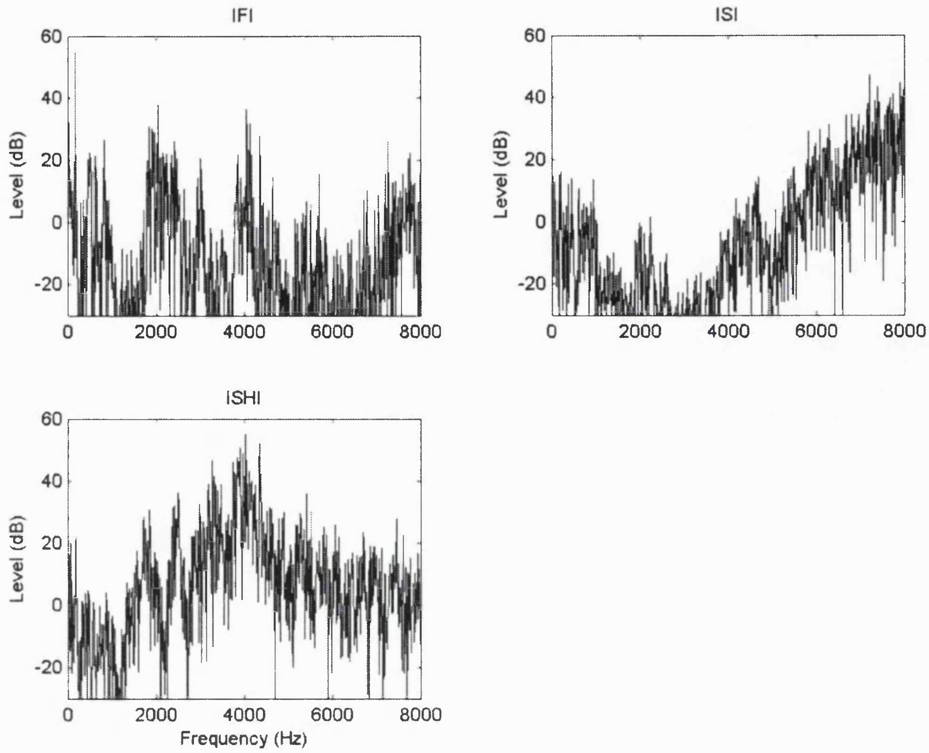
## 12.1 The Original natural stimuli

The measurements made on these stimuli are explained and summarised in Table 1 of chapter five. It is important to note that the stimuli used in the experiment were produced from the average measurements of 15 tokens for each fricative and each five were in one of three vowel environments (/i/, /a/ or /u/)

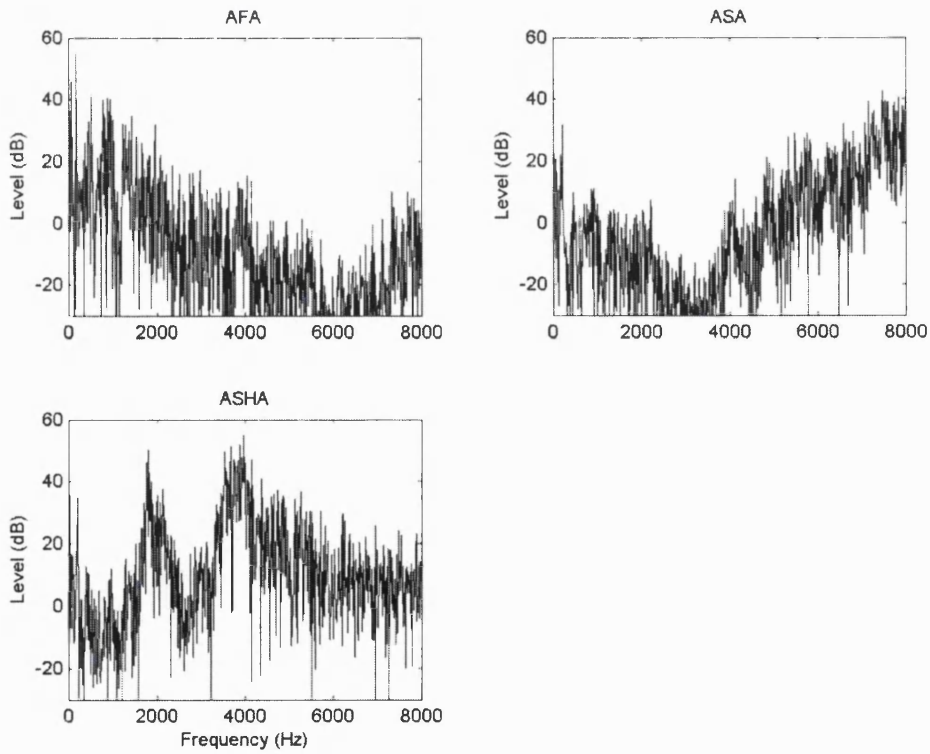
Fricatives in the /u/ environment



Fricatives in the /i/ environment




Fricatives in the /a/ environment



## 12.2 Processed Stimuli

The spectra below are produced from the stimuli created for testing identification of simplified fricatives. An explanation of how they were created is given in chapter five.