Speech, Hearing and Language: work in progress

Volume 14

EFFECTS OF THE NUMBER OF SPEECH-BANDS AND ENVELOPE SMOOTHING CONDITION ON THE ABILITY TO IDENTIFY INTONATIONAL PATTERNS THROUGH A SIMULATED COCHLEAR IMPLANT SPEECH PROCESSOR.

Matthew SMITH and Andrew FAULKNER

Department of Phonetics and Linguistics
UNIVERSITY COLLEGE LONDON

EFFECTS OF THE NUMBER OF SPEECH-BANDS AND ENVELOPE SMOOTHING CONDITION ON THE ABILITY TO IDENTIFY INTONATIONAL PATTERNS THROUGH A SIMULATED COCHLEAR IMPLANT SPEECH PROCESSOR.

Matthew SMITH and Andrew FAULKNER

Abstract

Though many investigations have been performed into the way that dividing the speech spectrum into a number of channels effects the intelligibility of speech, there appear to be few studies that have examined the effects of manipulating the number of channels on the ability to recognise and identify intonational patterns in speech. Secondly, many studies have demonstrated that voice fundamental frequency plays a significant role in the perception of speech intonation patterns; available evidence also suggests that changes in signal processing (in terms of simplifying the input waveform) could improve the perception of voice pitch changes for users of hearing aids and cochlear implants. In this study the ability to identify five intonational modes was assessed using sentences produced by a single male talker. The identification scores from seven normally-hearing subjects were examined with speech presented through 1, 4, 8, and 16-band simulations of cochlear implant signal processors (noise-band vocoders). Additionally, three different envelope conditions were utilised with each processor; this allowed the salience of temporal envelope cues to fundamental frequency to be manipulated and the consequent effect on identification performance to be examined. Significant improvements in identification ability were observed with both increased salience of temporal envelope cues to fundamental frequency and with increasing number of speech-bands (although the improvement in scores from the 8-band condition to the 16-band condition was not statistically significant). However, the level of identification improvement was found to vary significantly with intonational mode, suggesting that the overall identification scores may have been a reflection of subjects' ability to disambiguate one or two intonational modes from the rest.

1. Introduction

Speech signals are continuous and rapidly changing acoustic events in time. The aim of this study was to investigate the effect of varying (a) the amount of spectral detail available, and (b) the salience of temporal envelope cues to voice fundamental frequency, upon subjects' ability to identify different intonational patterns in speech.

For a listener, the ability to perceive the intonational features in speech is vital for the successful decoding of both linguistic and affective components of a spoken message. Regarding the latter, Cosmides (1983) claims that the communication of emotion expresses the speaker's intentions; the speaker's "attitude toward the actions, state of being, or persons represented....[in] the sentence's semantic structure embodies his or her behavioural intentions." (p.875). Prosodic aspects of communication, like intonation, are essential to understanding what a speaker intends to do about the state of affairs represented in an utterance. For example, a semantic or syntactic analysis of the phrase "I'll do it" provides no information regarding how the speaker feels about doing it. A more complete interpretation of the intended message of the speaker who says "I'll do it" in a tired, resigned tone of voice may well be "I'll do it, but I don't want to" (Cosmides, 1983). Such insight is likely to be available to the listener through their ability to interpret the prosodic characteristics of the utterance. Similarly, Grant and Walden (1996) and Pell (2001) assert that the basic function of prosody in speech is to provide information about lexical, grammatical and emotional aspects of the spoken message.

Reviewing studies of the expression of emotion in speech, Frick (1985) suggests that all prosodic features – stress, rhythm and intonation – are utilised in the expression of emotion while Cosmides (1983) points out that researchers interested in the acoustic expression of emotion usually assume that different individuals express the same emotions in similar ways. She reports agreement among many studies demonstrating that different individuals express particular emotions through standard formations of acoustic cues. Listeners must decode these cues in order to arrive at accurate judgements of speaker affect (Friend & Farrar, 1994). In his review Frick reports that increased pitch and loudness are used to signal the emotions anger, happiness and confidence, whereas low pitch and loudness levels are used to signal grief and boredom. Intonational contours – patterns of pitch and loudness over time – are also significant in distinguishing speaker emotion. The perception of these contours is also crucial in determining linguistic distinctions. For instance, in English, a rise in pitch at the end of a sentence can transform a statement into a question, although this "final rise" is generally associated with yes-no rather than wh-questions (McRoberts, Studdert-Kennedy & Shankweiler, 1995).

The design of this present study was significantly influenced by Lieberman and Michaels' paper of 1962. They used speech synthesis techniques to synthesize stimuli that differed from one another with respect to a single acoustic parameter present in normal speech and so examined the contributions of voice fundamental frequency (F0) and amplitude to the transmission of the emotional content in normal speech. The original amplitude and fundamental frequency information was extracted from sentences uttered in a variety of simulated emotional modes. By asking a group of subjects to listen to the stimuli and categorize them in terms of emotional mode, Lieberman and Michaels made deductions about the pertinence of particular acoustic parameters to the observed differences in the listeners' judgements regarding the stimuli. They found that both gross changes and the fine temporal structure of the fundamental frequency information were crucial to the correct identification of the original simulated emotion, but that the speech envelope amplitude contributed less to identification. However, it should be noted that the "emotional" categories, to which the subjects had to assign each sentence heard, might have been better described as "intonational" categories; the modes to be identified included not only ones representing speaker affect (emotion) such as fear, happiness, boredom and disbelief, but also modes where intonation is used for linguistic purposes, for instance in signalling the difference between a question and a statement. This is not a distinction made by Lieberman and Michaels, at least not overtly. It is important to mention this because, in the context of this author's study, when further reference is made to their paper, "emotion" and "emotional mode" will be used in the sense that Lieberman and Michaels used the terms. Elsewhere in this report, these terms can be assumed to refer to speaker affect only. In their study sentences were presented in the following emotional modes: (1) a bored statement, (2) a confidential communication, (3) a question expressing disbelief or doubt, (4) a message expressing fear, (5) a message expressing happiness, (6) an objective question, (7) an objective statement, and (8) a pompous statement.

One of the clearest findings of their investigation was that by removing spectral information from the speech stimuli, identification of emotional mode fell from 85% with unprocessed speech to 44% when only fundamental frequency information was presented. They did not, however, examine in detail how manipulating the amount of spectral detail available in processed speech affected the ability to correctly identify different modes; the more complete the spectral detail in a processed speech signal, the better will be the acoustic description of the sound in terms of the amount of energy at each frequency.

The relatively modest degree of spectral resolution available through cochlear implants is one of the factors limiting their effectiveness for users (Faulkner, Rosen & Wilkinson, 2001). Generally, cochlear implants make use of between 4 and 24 electrodes (Dorman, Loizou, Fitzke & Tu, 1998) and although it would seem reasonable to assume that speech intelligibility would improve with increasing number of active channels – and consequently increased spectral detail – the results of many studies suggest that the relationship is not straightforward. Eisenberg, Shannon, Martinez, Wygonski and Boothroyd (2000) report that asymptotic levels of performance have been achieved by adult cochlear implant users when listening to speech in quiet delivered through as few as four to six channels. Friesen, Shannon, Baskent & Wang (2001) quote studies that demonstrate an asymptote for speech in noise of four to seven channels. For speech in the quiet, similar results have been obtained with normally-hearing listeners using simulations of cochlear implant signal processors. Shannon, Zeng, Kamath, Wygonski and Ekelid (1995) reported that sentences in quiet could be understood with greater than 90% accuracy when just four channels were presented in this way. A very similar level of accuracy was reported by Loizou, Dorman and Tu (1999) with five channels (although the stimuli in this case had been produced by multiple speakers), and Dorman et al. (1998) found subjects reaching asymptotic performance, in quiet, with five channels of stimulation. Normally-hearing listeners, unlike the implant users studies by Friesen et al., are however able to be make use of increased spectral resolution beyond 8 channels when listening to speech in noise (Freisen et al., 2001; Fu et al., 1998).

In the studies referred to above, the primary focus was the effect upon speech intelligibility of varying the number of spectral channels; performance on vowel, consonant, word and sentence recognition tests commonly being measured. For Loizou et al. (1999) these results provide overwhelming evidence that the fine spectral detail present in naturally produced utterances is not vital for speech recognition. However, although Grant and Walden (1996) reported on listeners' ability to identify intonational patterns through narrow non-overlapping filters whose centre-frequencies spanned the speech-frequency range – their aim being to examine the extent to which prosodic information could be extracted from different spectral regions – the author of the present report could find no studies that examined the effect of varying the number of channels – and thus the degree of spectral detail – upon ability to identify intonational patterns. Therefore, one of the primary aims of this study was to examine this effect to see if the addition of more spectral channels would result in improved intonation identification. Grant and Walden (1996) presented evidence to suggest that information regarding intonation may be present across the speech spectrum, though they found that intonation patterns were better identified in low-frequency filtered speech than in high-frequency filtered speech. That they found that it is possible to extract the acoustic cues to intonation from anywhere in the speech spectrum is not completely unexpected because of the temporal nature of many prosodic cues. In manipulating the degree of spectral detail available to listeners in the present study, the aim was also to preserve the normal temporal variations of the sentence stimuli as much as possible. The approach taken was to test normally-hearing listeners in a task similar to that used by Lieberman and Michaels (1962), but using simulations of cochlear implant signal processors. To implement this technique, the time-intensity envelope is extracted from band-pass filtered speech and used to modulate noise bands; a more detailed explanation of this process is given below. Worthy of note is that, while systematically manipulating spectral information, this type of processing has been shown to preserve the normal temporal variations of speech; within-band spectral information is discarded, and only time-varying between-band level differences are available to signal spectral structure (Faulkner et al., 2001).

A listener's ability to decode the intonational characteristics of speech is dependent on the accurate perception, resolution, and tracking of features of the voice fundamental frequency of the speaker (Richardson, Busby, Blamey & Clark, 1998). Speakers use fundamental frequency to convey linguistic distinctions and also paralinguistic information, such as the affective state of the speaker. The findings of Lieberman and Michaels (1962) have already been mentioned, but other researchers have examined these effects also. Williams and Stevens (1972) noted that relative pitch, the pitch of an utterance in relation to a speaker's normal pitch range, is probably a better cue for the expression of emotion than absolute pitch. They also concluded that the fundamental frequency contour over time is the aspect of the speech signal that appears to provide the clearest indication of a speaker's affective state. In another investigation into the independent function of intonation contour type, fundamental frequency range, and voice quality, a wide fundamental frequency range, as well as intonation contours with an upward trend, were observed to signal arousal, annoyance and involvement (Ladd, Silverman, Tolkmitt, Bergmann & Scherer, 1985). Voice quality and fundamental frequency range were found to be responsible for 93% of the variance in judgements of speaker affect from digitised speech stimuli. Protopapas and Lieberman (1997), although finding no correlation between fundamental frequency range and perceived emotional stress, did find that the latter correlated highly with measures of the mean and maximum fundamental frequency of utterances. Friend and Farrar (1994) asked subjects to rate the level of excitement, happiness, and anger of speech stimuli under three different content-masking procedures. The fact that subjects performed reasonably well across the procedures was attributed by the authors to fundamental frequency cues, since these were preserved in each procedure. Turning from speaker affect to linguistic context, Pell and Baum (1997), working with patients with unilateral brain damage, found that mean fundamental frequency was significant in the discrimination of the linguistic categories "declarative", "interrogative" and "imperative". The importance of conveying voice pitch information through hearing aids to the hearing impaired has also been investigated. Aids that provide access to variations in voice fundamental frequency can significantly improve lip-reading ability (Rosen, Walliker, Brimacombe & Edgerton, 1989). Although previous studies with the House/3M single-channel cochlear implant had highlighted concerns that single-channel cochlear implant users may have difficulty discriminating prosodic cues, in their study Rosen et al. (1989) found that the House/3M device did provide important cues to intonation, suggesting that such information could be readily signalled by temporal features even in a single channel.

The findings of the various authors mentioned above and at the start of this report illustrate the fact that, taken together, the acoustic parameters connected to fundamental frequency play a significant role when it comes to perceiving different intonational patterns, both in “affective” speech and when intonation is used for linguistic purposes. This holds for both normally-hearing and hearing-impaired listeners, with evidence available that, for the latter group, changes in signal processing (in terms of simplifying the input waveform) could improve the perception of voice pitch changes (Rosen et al., 1989). Hence the second aim of this study was to examine the ability to identify different intonational modes as a function of the salience of temporal envelope cues to fundamental frequency – under each of the band-number conditions previously described – the hypothesis being that increased salience would result in improved identification scores. Furthermore, might the significance of acoustic parameters related to F0 be so important to the identification of intonational mode, that the effect of the salience of temporally encoded fundamental frequency information might be seen across conditions of processed speech that vary quite considerably with respect to the amount of spectral detail present?

Despite being the main concern of Lieberman and Michaels, the contribution of vocal pitch perturbations – irregularities in F0 excitation rate – to the transmission of “emotional mode” information was not addressed in this study. The concern of the present study was with the effect of manipulating the salience of fundamental frequency cues en masse rather than separately manipulating F0 range, maximum F0, F0 perturbations, and so forth. Nevertheless, consideration will be given to these features in the discussion at the end of this report.

Nor was the purpose of this study to investigate the influence of spectral detail and F0 salience on subjects’ ability to distinguish, specifically, speaker affect. Rather it was to examine the influence of these factors on subjects’ ability to distinguish broadly different intonational patterns. Consequently, a mixture of “affect” and “linguistic” intonational modes were used. These were: (1) a bored statement, (2) an objective question, (3) an objective statement, (4) a message expressing excitement or happiness, and (5) a question expressing disbelief or doubt. Clearly, this is a subset of those modes used by Lieberman and Michaels. These five modes were deemed adequate for the purposes of this investigation because it was felt that they were particularly distinguishable styles in normal speech to normally-hearing listeners, more so perhaps than “a confidential communication” and “a pompous statement”. Also, from a practical standpoint, they are likely to occur more often than the mode “fear” in everyday conversation. The recording of the sentences was performed by the author, and these five modes were also found to be the most readily and consistently reproducible. It was mentioned at the beginning of this report that the “final rise”, which distinguishes questions from statements in English, generally only occurs with yes-no questions. In recording the stimuli, the author attempted to apply this final rise to all sentences in the mode “an objective question” in order to make as clear as possible the distinction between this mode and the mode “an objective statement”.

It could be argued, in the context of the production of stimulus materials, that recourse to simulated intonational modes, particularly the affect modes “excitement/happiness”, “bored”, and “disbelief/doubt”, invites stereotyped or theatrical productions from the individual or individuals recording the stimuli (Cosmides, 1983). It has proved complicated though to collect naturalistic exemplars of vocal emotion while maintaining control of linguistic variables and recording quality (Pell, 2001). Frick (1985) reviewed studies in which stimuli expressing emotion had been produced by asking a speaker to produce that emotion, by inducing the emotion in the speaker before speaking, and by having judges rate stimuli as good or bad examples of a particular emotion. He reported that no differences between the three methods had been found. In an attempt to ensure that the stimuli in this study were good and identifiable examples of the five intonational modes, the scores of judges on two listening tasks were used to select the final stimuli (best exemplars) from a larger number of examples recorded by the author.

As already mentioned, the approach taken in this report was to test normally-hearing listeners with sentences processed through simulations of cochlear implant signal processors. Many sources of variability inherent in the testing of cochlear implant patients are eliminated by using normally-hearing listeners. These include differences between users in electrical stimulation strategy, electrode positioning, and cell body survival in the spiral ganglion. Of course, simulations do have their drawbacks; for instance they cannot replicate the dynamics of the current spread of an implant in the cochlea (Dorman et al., 1998). Even so, simulations with normally-hearing listeners can be considered representative of the optimal performance of users of cochlear implants when the spectral resolution of a simulation is comparable with the number of effective channels of electrical stimulation (Dorman & Loizou, 1998).

One of the primary aims of this study was to examine the ability to identify intonational modes as a function of the number of bands (or channels) through which the speech signal is presented. On this occasion signals were analysed (by band-pass filtering the signals into n bands, where n equalled 1, 4, 8 and 16), rectified and smoothed. The computed energy of the envelope in each band was used to modulate noise bands with widths corresponding to those of the analysis filters. These signals were then output to the listeners. The second aim of the study was to examine the effect of the salience of temporal cues to fundamental frequency on the ability to identify intonational mode. To this end, stimuli were synthesized which modified the salience of modulation envelope at those modulation rates that match voice fundamental frequency (approximately 50 to 350 Hz). This was achieved using rectification and low-pass smoothing filters. In conjunction with rectification, smoothing gives a good representation of the time-varying size of a signal coming through a given band-pass filter (Rosen and Howell, 1991). In this way, for the speech stimuli used in this study, three envelope conditions were produced, the effects of which could be examined for each of the band-number conditions already mentioned.

(1) a 400 Hz envelope condition, where envelope modulations in the region of the fundamental frequency were preserved, while faster changes were smoothed.

(2) a 32 Hz envelope condition, where envelope modulations faster than 32 Hz were smoothed, resulting in reduced salience of the fundamental frequency (relative to the first condition).

(3) an “enhanced” envelope condition where a second component, reflecting only envelope modulations in the region of the fundamental frequency, was added to the 400 Hz envelope. This served to create a high salience pitch envelope condition with extra emphasis given to modulation rates in the region of voice fundamental frequency.

It is essential to note that the sentences used in this study – the frameworks for the conveyance of the intonational features – were considered to be emotionally neutral. The importance of this, of course, is that listeners' judgements should not be influenced by the semantic content of the message they hear. The eight sentences considered neutral by Lieberman and Michaels (1962) were used on this occasion.

2. Design

2.1 Speech Materials

In this study, a recording was made by a native male speaker of British English (the author) of eight short sentences, neutral with regard to their affective content. The sentences were those used by Lieberman and Michaels (1962).

- They have bought a new car.
- His friend came home by train.
- You have seen my new house.
- John found him at the phone.
- The lamp stood on the desk.
- They parked near the street light.
- We talked for a long time.
- He will work hard next term.

The recording took place in an anechoic recording room, the speaker reading each sentence with appropriate vocal modifications so that it could be identified as belonging to one of the following five intonational modes: (1) a bored statement, (2) an objective question, (3) an objective statement, (4) a message expressing excitement/happiness, (5) a question expressing disbelief or doubt. Each sentence was read three times in each intonational mode, the product being 120 sentence productions in total. The fifteen repetitions of each sentence were then categorized by five final-year speech and language therapy students, so that the "best" utterances – the most identifiable example from the three repetitions recorded, of each of the five intonational modes for each of the eight sentences – could be selected. Identification of these "best exemplars" was based on the results of two listening tasks performed by the five students. In the first task the fifteen repetitions of each sentence were placed in random order. All 120 sentences were then played to the listeners who were required to make a forced choice regarding the intonational style of each sentence presented. Again, for the second listening task, all 120 sentences were played to the listeners. On this occasion the sentences were ordered according to intonational style, so the 24 sentences read in a "bored" manner were followed by the 24 sentences read in the style of an objective question, and so forth. The order of presentation of the sentences within each 24-sentence block was randomised. For this task the listeners used a rating scale, from 1 = very poor, to 5 = very good, to indicate how good an example of the intended intonational style they thought each sentence.

The forced choice task was presented to the listeners before the rating task in order to minimise any learning effect. Such an effect might arise from exposure to the rating task and consequently interfere with the forced choice task if this were presented second because rating necessitated the listeners being aware of the intonational style of the sentences to which they were listening. Results from the two tasks were used to select the most identifiable example of each sentence in each intonational mode;

rating scores were considered first of all but when these failed to identify a single best exemplar, because of tied scores, the results from the forced choice task were also taken into account. In two instances further tied scores meant that a final decision regarding selection had to be taken by the author. In this way 40 sentences were selected – a recording of each of the eight sentences in each of the five intonational modes.

2.2 Preparation of Test Stimuli

Noise-excited vocoders similar to those described by Shannon et al. (1995) were used in the speech processing of the sentences. One purpose of the study was to examine ability to identify intonational style as a function of spectral resolution – number of spectral bands – so processors were designed with 1,4,8, and 16 bands, with each processor encompassing an overall frequency range of 100 to 5000 Hz. Processing was implemented in MATLAB. The stages of processing in each band comprised an analysis filter, half-wave rectification, envelope smoothing, modulation of a white noise by the envelope, and an output filter that was identical to the analysis filter for the same band. Finally, the outputs of each band were summed together. Each channel of the processor received speech as input, without pre-emphasis. Cross-over and centre frequencies for both the analysis and output filters were calculated using an equation (and its inverse) relating position on the basilar membrane – assuming it to be 35mm in length – to characteristic frequency, based on the cochlear tonotopic formulae of Greenwood (1990):

$$frequency = 165.4\left(10^{0.06x} - 1\right)$$

$$x = \frac{1}{0.06}\log\left(\frac{frequency}{165.4} + 1\right)$$

where x represents the distance along the basilar membrane (in mm) from the apex and frequency is in Hz. The channel filter –3 dB cut-off frequencies are shown in Table 1 for each of the 1, 4, 8 and 16 band processors. The filter centre frequencies were mid-way between adjacent cut-offs.

The second purpose of the study was to examine the ability to identify intonational style as a function of the salience of temporal envelope cues to fundamental frequency, so for each of the 1,4,8, and 16-band processors three different envelope conditions were utilised – producing twelve speech processed conditions. In the first envelope condition an envelope-smoothing filter with a low-pass cut-off frequency of 32 Hz was used, effectively eliminating F0. In the second envelope condition an envelope-smoothing filter with a low-pass cut-off frequency of 400 Hz was used in order to preserve F0. In order to enhance the salience of F0 a second component reflecting modulations only between 50Hz and 400 Hz was added to the 400 Hz envelope used in the second condition above. These higher rate components were added after multiplication by 100 (40dB). This third envelope condition was labelled "the enhanced condition". A summary of the twelve conditions of processed speech appears in Table 2 below.

	Processor				
	16-band	*8-band*	*4-band*	*1-band*	
-3dB cut-off frequencies (Hz)	100	100	100	100	1
	154	219	392	5000	2
	219	392	1005		3
	298	642	2294		4
	392	1005	5000		5
	507	1531			6
	642	2294			7
	807	3398			8
	1005	5000			9
	1244				10
	1531				11
	1877				12
	2294				13

Table 1: *–3dB cut-off frequencies for each of the 1,4,8, and 16 band processors.*

Off-line processing was executed at a 44.1 kHz sample rate. Analysis filters in the off-line processing were Butterworth IIR designs with 3 orders per upper and lower side. The responses of adjacent filters crossed 3 dB down from the pass-band peak. Envelope smoothing used 2nd-order low-pass Butterworth filters.

Condition n$^{o.}$	***Code***	***Number of Bands***	***Envelope***
1	B1:E32	1	32 Hz cut-off
2	B1:E400	1	400 Hz cut-off
3	B1:Eenh	1	Enhanced
4	B4:E32	4	32 Hz cut-off
5	B4:E400	4	400 Hz cut-off
6	B4:Eenh	4	Enhanced
7	B8:E32	8	32 Hz cut-off
8	B8:E400	8	400 Hz cut-off
9	B8:Eenh	8	Enhanced
10	B16:E32	16	32 Hz cut-off
11	B16:E400	16	400 Hz cut-off
12	B16:Eenh	16	Enhanced

Table 2: *Speech processing conditions.*

2.3 Procedure

Testing took place in a sound-isolated room. Processed and unprocessed speech was presented to subjects diotically through AKG K240DF headphones at a fixed comfortable listening level. Each subject was required to listen to the forty sentences selected from the listening tasks in each of the twelve speech-processed conditions in turn. For each condition, the sentences were presented automatically via a computer, and always in a random order. For each sentence presented, the subject was required to make a forced choice about which of the five intonational styles they believed they had heard: (1) a bored statement, (2) an objective question, (3) an objective statement, (4) a message expressing excitement/happiness, or (5) a question expressing disbelief or doubt. Five boxes on the computer screen in front of the subject were labelled for each of the five intonational styles, and the subject was required to respond by mouse-clicking on the relevant box. Only once the subject had made a response would the computer program play the next sentence; however, before the commencement of testing, the subject was instructed to not spend too long considering what they heard but rather to go with their instinctive response. Correct and incorrect categorizations were recorded automatically by the computer.

The subjects for this study were seven native English-speaking adults whose audiometric thresholds were tested and found to be within normal limits between 125 and 4000 Hz. All subjects were presented with the 40 sentences as unprocessed speech as the first condition so that their ability to identify the intonational styles in ideal conditions could be used as a benchmark, it also served to familiarise them with the test procedures. Subjects then moved on to the twelve processed speech conditions, the order of presentation of the conditions being randomly generated for each individual. Subjects were informed when they were moving on to a new condition. Finally the unprocessed speech condition was presented again. To minimise fatigue effects, the subjects undertook half the testing on one day and half on a second occasion (within a week of the first).

3. Results

A simple view of the data is shown in Figure 1. The proportion of correct intonational mode identifications, taken across all seven subjects, is illustrated for the unprocessed speech condition and each of the twelve speech-processed conditions. This data illustration, for each given condition, incorporates one mean score for each subject. The mean scores, taken across all subjects, are presented in Appendix 1 (Table A1).

Without separating the scores for each intonational mode, the data above would appear to suggest that correct identifications were being made at better than chance level (20% correct) in all speech-processed conditions, mean scores ranging from 29% in the 1-band 400 Hz condition to 51% in the 16-band enhanced condition. Despite this, scores for speech-processed conditions never reached the level of identification attained with unprocessed speech. An analysis of variance was performed on the data; using the Huynh-Feldt corrected F test, a main effect due to the number of bands of speech was revealed ($F[2.23,13.4] = 10.6$, $p<0.005$). Similarly, a main effect due to envelope condition was also revealed ($F[1.21,7.23] = 14.2$, $p<0.01$), however there was found to be no interaction of number of bands with envelope condition ($F[6,36] = 1.47$, $p=0.216$).

At the outset of the study it was predicted that each increase in the number of bands, and each step in envelope condition, from 32 Hz to 400 Hz to 400 Hz enhanced, should result in improved ability to identify intonational mode. An a priori contrast test was used to examine the significance of these predicted differences between pairs of levels within each factor. A summary of the analysis can be seen in Table 3 below.

Taking an alpha of 0.05, the test indicated that identification performance with 8-band speech (46.7% correct) was significantly better than with 4-band speech (41.7%), which in turn was significantly better than with 1-band speech (33.2%). Performance, however, appeared to reach asymptote with 8-band speech, results for the 8-band and 16-band (49.1% correct) speech-conditions not being significantly different. Turning to the levels within the envelope factor: identification performance in the 400 Hz enhanced condition (47.6% correct) was significantly better than in the 400 Hz condition (42.5%). The difference between performance levels in the latter condition and in the 32 Hz condition (38% correct), where F0 had effectively been eliminated, was also significant, performance in the 400 Hz condition being better.

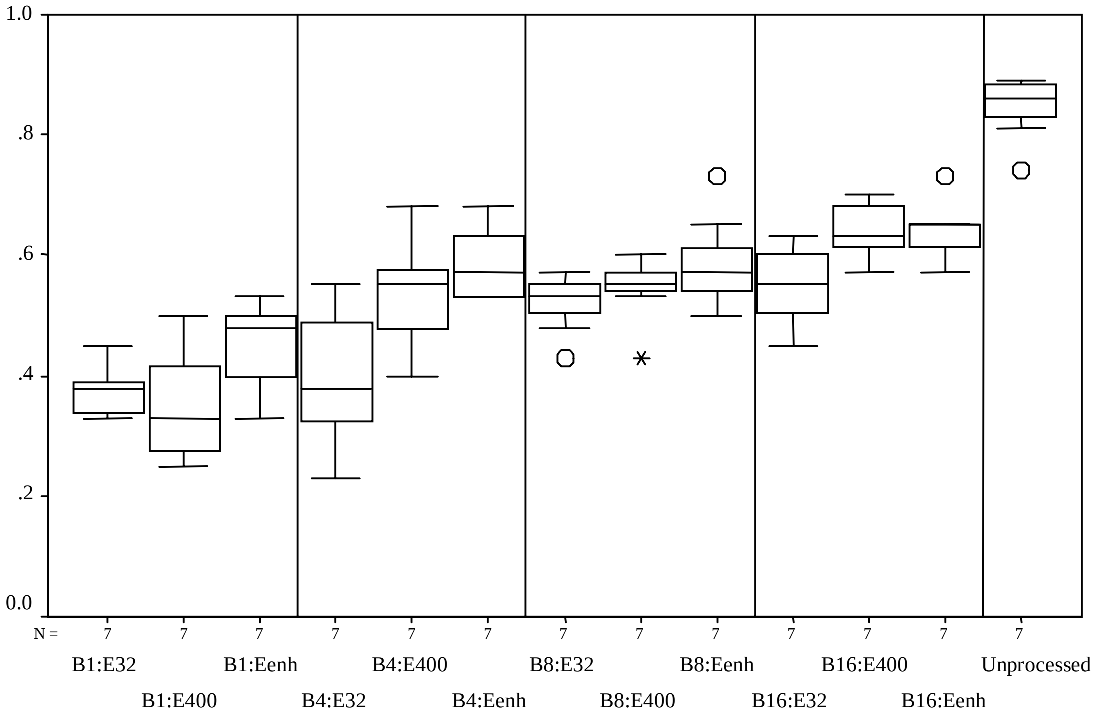

Figure 1: *Box and whisker plots, across subjects, of proportion of correct intonational mode identifications for each condition tested. The box indicates the inter-quartile range of values obtained, each based on seven scores, one from each subject. The median is represented by the solid horizontal line. The whiskers indicate the range of measurements, except for points more than 1.5 box lengths (indicated by 'o') or 3 box lengths (indicated by '*'). Refer to Table 2 for a key to the condition codes*

The descriptive statistics presented in Table A1 (Appendix 2) suggest that, for each band condition, increasing the salience of temporal envelope cues to F0 resulted in improved identification scores. This is particularly noticeable with the 4-band speech processor where correct identification rose from 32% in the 32 Hz condition to 49% in the 400 Hz enhanced condition. However, the improvement was less marked with the 16-band speech processor where, for the same envelope conditions, the improvement was from 46% to 51% only.

Factor	Comparison	P
Number of Bands	1 band vs. 4 bands	0.028
	4 bands vs. 8 bands	0.042
	8 bands vs. 16 bands	0.287
Envelope	32 Hz vs. 400 Hz	0.001
	400 Hz. vs. 400 Hz enhanced	0.037

Table 3: *Significance of predicted differences between pairs of levels within each factor.*

The data was further analysed to see if certain intonational modes were identified more successfully than others. Table 4 shows the proportion of correct identifications for each intonational mode, taken across all subjects. Although the most successfully identified mode – "question expressing disbelief/doubt" – was identified correctly on nearly 3 out of 4 presentations, the least successfully identified mode – "an objective statement" – was correctly identified on only 1 in 7 presentations. An analysis of variance using Huynh-Feldt adjustments found the effect of intonational mode to be highly significant ($F[3.08,18.5] = 20.1$, $p<0.001$). Furthermore, a pairwise comparison of the identification scores for each intonational mode showed that identification of each of the modes "a bored statement", "message expressing excitement/happiness" and "question expressing disbelief/doubt", was significantly better than identification of the mode "an objective statement". Also, identification of the mode "question expressing disbelief/doubt" was significantly better than that of the mode "a bored statement".

Intonational mode	Bored statement	Objective question	Objective statement	Message expressing excitement/ happiness	Question expressing disbelief/doubt
Mean	.380	.301	.142	.562	.749

Table 4: *Proportion of correct identifications for each intonational mode. Means are based on the responses of all seven subjects.*

Appendix 2, Table A2 presents the proportion of correct identifications in each speech-processed condition, for each intonational mode separately. A further Huynh-Feldt adjusted F-test revealed a significant interaction between intonational mode and the number of speech-bands ($F[8.48,50.9] = 11.1$, $p<0.001$) and also between intonational mode and envelope condition ($F[8,48] = 5.60$, $p<0.001$). A 3-way interaction between intonational mode, number of speech-bands and envelope condition was not found. The existence of these interactions indicates that across the intonational modes different degrees of improved identification occurred as the number of speech-bands was increased and with the shift from one envelope condition to another (Figure 2). For instance, as the number of speech-bands was increased from 1-band to 16-bands, identification of the mode “question expressing disbelief/doubt” improved from 38.9% to 96.5%, however identification of “a bored statement” fell from 39.5% to 31.9%. Similarly, across the envelope conditions, going from 32 Hz (E32) to enhanced (Eenh), identification of most modes increased only slightly (or fell slightly in the case of the mode “message expressing excitement/happiness), whereas identification of “a bored statement” improved from 20.8% to 54.3%.

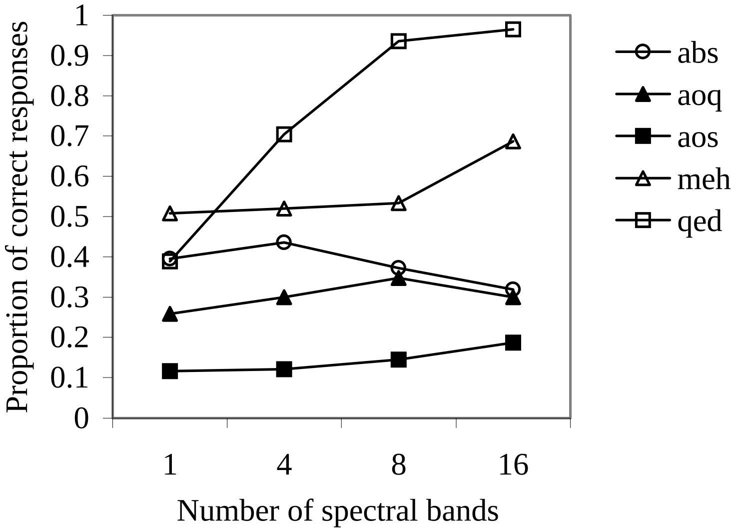

Figure 2(a): *Proportion of correct identifications of each intonational mode, across subjects, for different numbers of bands (abs: a bored statement, aoq: an objective question, aos: an objective statement, meh: message expressing excitement/happiness, qed: question expressing disbelief/doubt).*

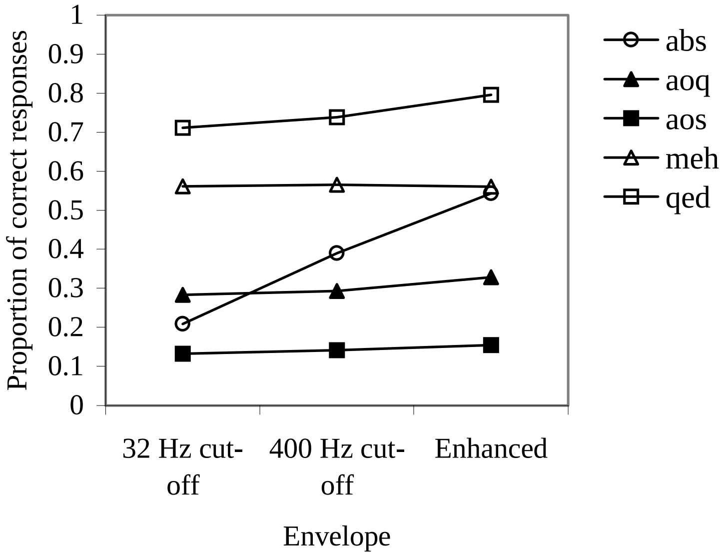

Figure 2(b): *Proportion of correct identifications of each intonational mode, across subjects, for different envelope conditions.*

A chi-square test was performed to check Lieberman and Michaels' (1962) assumption that the eight sentences used in this study were neutral with respect to intonational/affective modality and thus should not significantly influence identification. If this were true, the total for correct identifications would be expected to be fairly evenly distributed between the eight sentences. Testing the hypothesis that each of the eight sentences appears in the list of correct identifications equally often, a value for χ^2 was obtained: $\chi^2 = 4.50$. The critical value for χ^2 for df = 7 at $p<0.05$ is 14.07, so Lieberman and Michaels' assertion is supported and the ability to correctly identify intonational mode in the present study can be considered to be independent of the sentences used.

Throughout the data analysis, simple measures of proportion of correct identifications were used. As these measures could be affected by participants having a bias of choice for a particular intonational mode, an unbiased measure – one of proportion information transfer – was also computed. A scatter-plot of information transfer against proportion correct displayed a fairly linear relationship with a high correlation ($r = 0.81$). Consequently, there would appear to be no reason to reject the use of the simple measure of "proportion of correct identification".

4. Discussion

4.1 Number of Speech Bands

Lieberman and Michaels' (1962) finding that correct identification of emotion dropped from 85% to 47% when spectral information was excluded from sentence stimuli, demonstrated the importance of spectral structure to the perception of emotion in speech. In the present study, increasing the number of bands (channels) of speech, and thus the degree of spectral detail available to the listener, improved the subjects' overall ability to identify the intonational patterns of the presented speech stimuli. Although identification performance improved with increasing number of speech bands, there also appeared to be a "slowing-down" in the degree of improvement across the four conditions; in the 1-band condition 33% of identifications were correct, in the 4-band condition 42%, in the 8-band condition 47%, and in the 16-band condition 49% correct. Statistically, the improvement demonstrated between the 8-band condition and the 16-band condition was not significant, even though the number of speech-bands had been doubled.

The general pattern of improvement is what one might have expected, since the spectral resolution of the stimulus signals increases with increasing number of speech-bands. The vocoder-like speech-processing technique used in this study produced stimuli where within-band spectral information was discarded, but where time-varying between-band level differences were still available to signal spectral structure. Of course, in the 1-band condition between-band comparisons were not available to the subjects and the acoustic energy in the sentence stimuli would have been spread across the entire frequency range of the filter-band from 100 Hz to 5000 Hz. Consequently, negligible information would have been available regarding the amount of energy at each frequency in the original sentence items. This raises the question as to why identification performance in the 1-band condition was even as high as 33% correct – significantly better than chance level at 20%.

With the 4-band speech condition performance rose to 42% correct; however, the lower and upper cut-off frequencies for the lowest-frequency band were 100 Hz and 392 Hz respectively. Consequently, all the energy related to frequencies in the region occupied by the fundamental frequency (F0) is likely to have been expressed in this first band. Therefore, it is doubtful that variability in spectrally-signalled fundamental frequency features, such as F0 contour, range, mean, and maximum would have been influential in the improved intonation identification observed with 4-band speech, since these features could only become salient when the frequency region occupied by F0 falls across more than one band, thus enabling between-band comparisons, as illustrated by Williams and Stevens (1972). As Grant and Walden (1996) demonstrated, it is possible to extract the necessary acoustic cues for intonation from anywhere in the speech spectrum, not just from the region occupied by the fundamental frequency; the greater spectral resolution resulting from having four speech-bands may have increased the subjects' access to these cues from higher frequency regions, thereby leading to improvement in identification. Frick (1985) reports that energy may be shifted to higher frequencies during emotion; specifically, the first formant is widened and shifted upward; for positive emotions the upward shift may be greater than for negative emotions. This may explain why in the 4-band condition, identifications of the modes "question expressing disbelief/doubt" (similar to surprise) and "message expressing excitement/happiness" were relatively high at 70% and 52% respectively, while identifications of the linguistic (non-affect) modes "objective statement" and "objective question" were significantly lower at 12% and 30%.

Access to cues to intonation in higher frequency regions is likely to have improved further in the 8-band and 16-band conditions and may be partly responsible for the improved identification scores in these conditions. Additionally, resolution of the frequency region occupied by the fundamental frequency would have been better in these conditions; from Figure 2 it can be seen that this region (approximately 50 to 350 Hz) fell across two filter-bands in the 8-band condition and across four bands in the 16-band condition. With increased spectral resolution of this region, subjects would have been able to combine "snap-shot" between-band comparisons and perceptions of the movement of energy between bands across the length of an utterance, giving an insight into features of the fundamental frequency such as F0 contour, F0 mean, and F0 range (it is doubtful whether 16-band resolution would have significantly improved perception of F0 maximum). Improved perception of these features may well have been responsible (at least in part) for the further increases in identification scores exhibited in the 8-band and 16-band conditions.

Now to return to the question of why identification performance in the 1-band condition was as high as it was. Rosen et al. (1989) point to the fact that spectral shape distinctions can be conveyed through the temporal microstructure of waveforms applied to a single channel. However, the temporal cues in the stimuli used in this study were only preserved up to modulation rates within the pass-band of the envelope-smoothing filter. In the 1-band/32 Hz envelope condition (B1:E32), where temporal microstructure related to F0 would have been absent, 31% of identifications were correct – still above chance level by some way. Perhaps more likely is that some gross temporal cues across the duration of the utterances, such as segment/interval duration, were salient enough in the condition B1:E32 to account for this unexpectedly high level of identification. An examination of the figures in Table A2 (Appendix 2) reveals that the level of identification in this condition was primarily influenced by the identification score for one particular mode: the mode "message expressing excitement/happiness" was identified correctly in 57% of presentations in condition B1:E32 (the next highest score was for the mode "question expressing disbelief/doubt" at 32%). Frick (1985) comments that happiness can be signalled by a faster rate of speech; although the sentences in this study were short (six syllables) in length and the author made a conscious effort, during recording, to produce utterances of even duration, it is possible that small differences in stimulus length may have been enough to disambiguate one or two of the intonational modes from the rest even when spectral resolution was at its poorest.

Another question that arises is why do the improvements in identification scores appear to level off across the levels of the band-number condition? As already mentioned, the improvement demonstrated between the 8-band condition and the 16-band condition – just 2% – was not significant, even though the number of speech-bands had been doubled. The answer to this is not entirely obvious; in the introduction it was observed that several studies with both normally-hearing listeners and cochlear implant users found that, with speech intelligibility tests in quiet, asymptotic levels of performance were reached when listening to speech through as few as four to seven channels (Eisenberg et al., 2000; Friesen et al., 2001; Shannon et al., 1995; Loizou et al., 1999; Dorman et al., 1998). According to Loizou et al. these results indicate that the fine spectral detail present in naturally-produced utterances is not vital for speech recognition; possibly this is true for the recognition of intonation as well. It may be enlightening to repeat the present study with more levels either side of the 8-band condition – maybe 6, 10 and 12 – and with individuals with cochlear implants, in order to further investigate this levelling-off in improvement.

4.2 Envelope Condition

Consideration will now be given to the effect on identification scores of manipulating the fundamental frequency salience of the sentence stimuli. Lieberman and Michaels (1962) reported that voice pitch information played an important role in the transmission of emotional mode information even when a complete spectral description of the speech material is available. In this study the salience of the region of the speech signals occupied by the fundamental frequency was manipulated using the three different envelope-smoothing conditions described in the Introduction and Design sections. Temporal cues were preserved up to modulation rates dependent on details of the envelope extraction. Although significant improvements in identification scores were observed across the three conditions, the improvements were not large. They amounted to a 10% increase in correct identifications from the 32 Hz envelope condition to the 400 Hz enhanced envelope condition – a rise from 38% correct to 48% correct. Nevertheless, the potential for improving voice pitch perception by signal processing (by simplifying the input waveform) is of crucial importance to the field of cochlear implant technology (Rosen et al., 1989); in this study the enhanced saliency condition was produced by adding a second component to the 400 Hz envelope, reflecting only modulations between 50 Hz and 400 Hz. This did produce a significant, if small, improvement in intonation pattern identification and, as a signal processing technique, is worth further investigation. Additionally, the absence of an interaction between the speech-band number and envelope conditions suggests that the improvements observed for both main effects were produced independently of each other.

4.3 Intonational Modes

Significant interactions were observed between intonational mode and the number of speech-bands, and between intonational mode and envelope condition, indicating that different degrees of improvement occurred across modes as the number of speech-bands was increased, and with the shift from one envelope condition to the next. Figure 2 (a) illustrates the increased identification performance observed with increasing number of speech-bands which can be attributed, in the main, to a large improvement in the identification of the mode "question expressing disbelief/doubt" from the 1-band to the 8-band condition (39% correct to 94% correct), and a significant improvement in the identification of the mode "message expressing excitement/happiness" between the 8-band and 16-band conditions (53% correct to 69% correct). For the remaining modes: "a bored statement", "an objective question", and "an objective statement", identification scores show little in the way of improvement; in fact identification of the mode "a bored statement" fell from a high in the 4-band condition (44% correct) to a low in the 16-band condition (32% correct). Further consideration of Figure 2 (a) suggests that increasing spectral resolution enabled the subjects to distinguish the two modes "question expressing disbelief/doubt" and "message expressing excitement/happiness" from the other three modes.

Demonstrating that a set of intonational patterns can be discriminated from one another better than would be expected by chance does not prove that all the patterns in the set are being reliably identified. Above chance discrimination might, in fact, only imply that one or two of the patterns can be reliably discriminated from the others. In this study the sentences in the mode "question expressing disbelief/doubt" were recorded with features similar to those of "surprise": a wide pitch range and a rising pitch contour overall (Ladd et al., 1985, Frick, 1985), while the sentences in the mode "message expressing excitement/happiness" were characterised by a relatively high voice pitch across the length of each utterance (see Pell, 2001, among others). However, in recording the two linguistic intonational modes – objective statement/objective question – the author's aim was to produce naturally sounding utterances (with lower voice pitch than the aforementioned modes) that differed in the terminal glide – a rise for "question", a fall for "statement" (Studdert-Kennedy & Hadding, 1973; Pell and Baum, 1997; McRoberts et al., 1997). Sentences in the affect mode "bored statement" were recorded in a fairly flat, low voice pitch. Looking at Figure 2 (a) we might hypothesise that as the number of speech-bands increased, the increased salience of fundamental frequency features, as well as of cues to intonation in higher frequency regions, enabled subjects to improve their identification performance. However, the increase in spectral resolution may have been sufficient to enable subjects to distinguish the two modes ("question expressing disbelief/doubt" and "message expressing excitement/happiness") that are characterised by relatively steep F0 contours, wide F0 ranges, and high mean F0 from the three modes characterised by more subtle F0 variability, and yet have been insufficient to disambiguate these three modes from each other. It is certainly possible that the F0 contours of the terminal glides of the modes objective statement/objective question, although distinctive enough to be discriminated in unprocessed speech, were too subtle to be perceived with any great accuracy in the processed speech conditions. It is also conceivable that the mode "bored statement" became a response of choice to some extent, which might explain its marginally superior identification when compared to the linguistic modes. The marked improvement of identification of the "bored statement" mode across the three envelope conditions (Figure 2 (b)), in the absence of significant improvements in the other modes, is less easily explained. It is possible that the increased salience of the fundamental frequency region simply made a low, even F0 contour disproportionately distinctive.

The fact that the subjects' overall level of identification performance did not rise much above 50% (51% in the 16-band 400 Hz enhanced envelope condition, B16:Eenh) can be traced to poor identification of the modes "objective question", "bored statement", and "objective statement", especially the latter (see Appendix 2, Table A2). Another explanation for the relatively low scores overall may be that rather than a perceptual problem, the difficulty for the subjects arose from the requirement to explicitly label contrasts that are only vaguely represented in our writing system (Rosen et al., 1989, Ladd et al., 1985). Possibly the most important weakness in this area of research is the absence of a widely accepted taxonomy of emotion and attitude. Further investigations continuing on from this study might benefit from focussing in a more concentrated manner on identification of intonation with different emotional modes or with different linguistic modes, rather than with a mix of the two as was used here.

In this investigation the subjects were not given any prior training with the stimulus materials used. Rosen et al. (1997) found that even a small amount of training enabled subjects to considerably improve their performance on connected discourse tracking with spectrally-shifted speech. As cochlear implants typically present spectral information to the wrong "place" along the basilar membrane, this was a significant finding. An improved version of the present study might make use of a period of training at the beginning, to see if subjects might then be able to adapt more successfully to the processed speech stimuli. A larger study with more subjects is likely to produce more reliable data; furthermore, the realistic nature of the investigation would benefit from the use of a larger number of stimuli, recorded by more speakers. In this study a single talker (the author) recorded the sentence stimuli, but as Loizou et al. (1999) point out, this overestimates the speech perception abilities of listeners in real-world situations where acoustic signal variability arises from differences in the size and shape of vocal tracts, differences in speaking rate, and differences in the way people use prosody in speech.

5. Acknowledgement

This report is based on Matthew Smith's BSc Speech Communication Project completed in 2002. We gratefully acknowledge the support of the Royal National Institute for the Deaf (RNID).

References

Cosmides, L. (1983). Invariances in the Acoustic Expression of Emotion During Speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 864-881.

Dorman, M.F., Loizou, P.C., Fitzke, J. & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear implant signal processors with 6-20 channels. Journal of the Acoustical Society of America, 104, 3583-3585.

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

Faulkner, A., Rosen, S. & Wilkinson, L. (2001). Effects of the Number of Channels and Speech-to-Noise Ratio on Rate of Connected Discourse Tracking Through a Simulated Cochlear Implant Speech Processor. Ear and Hearing, 22, 431-438.

Frick, R.W. (1985). Communicating Emotion: The Role of Prosodic Features. Psychological Bulletin, 97, 412-429.

Friend, M. & Farrar, M.J. (1994). A comparison of content-masking procedures for obtaining judgements of discrete affective states. Journal of the Acoustical Society of America, 96, 1283-1290.

Friesen, L.M., Shannon, R.V., Baskent, D. & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. Journal of the Acoustical Society of America, 110, 1150-1163.

Grant, K.W. & Walden, B.E. (1996). Spectral Distribution of Prosodic Information. Journal of Speech and Hearing Research, 39, 228-238.

Greenwood, D.D. (1990). A cochlear frequency-position function for several species – 29 years later. Journal of the Acoustical Society of America, 87, 2592-2605.

Ladd, D.R., Silverman, K.E.A., Tolkmitt, F., Bergmann, G. & Scherer, K.R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. Journal of the Acoustical Society of America, 78, 435-444.

Lieberman, P. & Michaels, S.B. (1962). Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech. Journal of the Acoustical Society of America, 34, 922-927.

Loizou, P.C., Dorman, M. & Tu, Z. (1999). On the number of channels needed to understand speech. Journal of the Acoustical Society of America, 106, 2097-2103.

McRoberts, G.W., Studdert-Kennedy, M. & Shankweiler, D.P. (1995). The role of fundamental frequency in signalling linguistic stress and affect: Evidence for a dissociation. Perception & Psychophysics, 57, 159-174.

Pell, M.D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. Journal of the Acoustical Society of America, 109, 1668-1680.

Pell, M.D. & Baum, S.R. (1997). Unilateral Brain Damage, Prosodic Comprehension Deficits, and the Acoustic Cues to Prosody. Brain and Language, 57, 195-214.

Protopapas, A. & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. Journal of the Acoustical Society of America, 101, 2267-2277.

Richardson, L.M., Busby, P.A., Blamey, P.J. & Clark, G.M. (1998). Studies of Prosody Perception by Cochlear Implant Patients. Audiology, 37, 231-245.

Rosen, S., Faulkner, A. & Wilkinson, L. (1997). Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants. Speech Hearing and Language: work in progress, 10, 1-15.

Rosen, S. & Howell, P. (1991). *Signals and Systems for Speech and Hearing.* London: Academic Press.

Rosen, S., Walliker, J., Brimacombe, J.A. & Edgerton, B.J. (1989). Prosodic and segmental aspects of speech perception with the House/3M single-channel implant. Journal of Speech and Hearing Research, 32, 93-111.

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

Studdert-Kennedy, M. & Hadding, K. (1973). Auditory and Linguistic Processes in the Perception of Intonation Contours. Language and Speech, 16, 293-313.

Williams, C.E. & Stevens, K.N. (1972). Emotions and Speech: Some Acoustical Correlates. Journal of the Acoustical Society of America, 52, 1238-1250.

Appendix 1: Statistical Data.

Table A1: *Proportion of correct intonational mode identifications under each speech processed condition. Means are based on the responses of all seven subjects.*

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.306	.065	7
B1:E400	.293	.113	7
B1:Eenh	.399	.166	7
B4:E32	.321	.094	7
B4:E400	.437	.090	7
B4:Eenh	.491	.087	7
B8:E32	.431	.054	7
B8:E400	.467	.067	7
B8:Eenh	.501	.067	7
B16:E32	.460	.061	7
B16:E400	.501	.076	7
B16:Eenh	.511	.083	7

Code	Number *of Bands*	Envelope
B1:E32	1	32 Hz cut-off
B1:E400	1	400 Hz cut-off
B1:Eenh	1	Enhanced
B4:E32	4	32 Hz cut-off
B4:E400	4	400 Hz cut-off
B4:Eenh	4	Enhanced
B8:E32	8	32 Hz cut-off
B8:E400	8	400 Hz cut-off
B8:Eenh	8	Enhanced
B16:E32	16	32 Hz cut-off
B16:E400	16	400 Hz cut-off
B16:Eenh	16	Enhanced

Table A2: *Proportion of correct identifications for each intonational mode (a to e) under each speech processed condition. Means are based on the responses of all seven subjects.*

(a) A Bored Statement

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.270	.198	7
B1:E400	.396	.266	7
B1:Eenh	.520	.349	7
B4:E32	.144	.112	7
B4:E400	.466	.212	7
B4:Eenh	.699	.278	7
B8:E32	.200	.097	7
B8:E400	.341	.223	7
B8:Eenh	.574	.189	7
B16:E32	.219	.119	7
B16:E400	.359	.254	7
B16:Eenh	.379	.192	7

(b) An Objective Question

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.270	.196	7
B1:E400	.216	.187	7
B1:Eenh	.289	.305	7
B4:E32	.216	.214	7
B4:E400	.341	.351	7
B4:Eenh	.341	.247	7
B8:E32	.323	.271	7
B8:E400	.360	.224	7
B8:Eenh	.359	.328	7
B16:E32	.323	.202	7
B16:E400	.253	.177	7
B16:Eenh	.324	.278	7

(c) An Objective Statement

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.093	.063	7
B1:E400	.074	.069	7
B1:Eenh	.180	.288	7
B4:E32	.146	.135	7
B4:E400	.090	.119	7
B4:Eenh	.127	.102	7
B8:E32	.073	.099	7
B8:E400	.183	.143	7
B8:Eenh	.180	.143	7
B16:E32	.216	.059	7
B16:E400	.217	.226	7
B16:Eenh	.129	.127	7

(d) A Message Expressing Excitement/Happiness

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.574	.189	7
B1:E400	.447	.249	7
B1:Eenh	.501	.192	7
B4:E32	.486	.183	7
B4:E400	.519	.253	7
B4:Eenh	.554	.175	7
B8:E32	.591	.200	7
B8:E400	.576	.202	7
B8:Eenh	.431	.161	7
B16:E32	.591	.248	7
B16:E400	.717	.140	7
B16:Eenh	.753	.143	7

(e) A Question Expressing Disbelief/Doubt

Descriptive Statistics			
	Mean	**Std.Deviation**	**N**
B1:E32	.323	.202	7
B1:E400	.341	.266	7
B1:Eenh	.503	.289	7
B4:E32	.609	.273	7
B4:E400	.770	.198	7
B4:Eenh	.734	.168	7
B8:E32	.966	.059	7
B8:E400	.877	.143	7
B8:Eenh	.966	.059	7
B16:E32	.947	.098	7
B16:E400	.964	.095	7
B16:Eenh	.983	.045	7