Speech, Hearing and Language: work in progress

Volume 11

Periodicity and pitch information in simulations of cochlear implant speech processing

Andrew FAULKNER, Stuart ROSEN and Clare SMITH

Department of Phonetics and Linguistics
UNIVERSITY COLLEGE LONDON

Periodicity and pitch information in simulations of cochlear implant speech processing

Andrew FAULKNER, Stuart ROSEN and Clare SMITH

Abstract

Pitch, periodicity and aperiodicity are regarded as important cues for the perception of speech. However, modern CIS cochlear implant speech processors, and recent simulations of these processors, provide no explicit representation of these factors. We have constructed four-channel vocoder processors that manipulate the representation of periodicity and pitch information, and examined the effects on the perception of speech and the ability to identify pitch glide direction.

A vocoder providing highly salient pitch and periodicity information used a pulse train source during voiced speech, and a noise source in the absence of voicing. The pulse train was controlled by voice fundamental frequency. A second condition provided a salient auditory contrast to periodicity but no pitch information, through the use of a fixed rate pulse source during voicing, and a noise source at other times. Further processing conditions were independent of input speech excitation. One such condition used a constant pulse train throughout, with neither periodicity nor pitch represented. Two further conditions used a noise source throughout. In one noise condition, the amplitude envelope extracted from each band was low-pass filtered at 32 Hz, eliminating pitch and periodicity cues from the envelope. In the second noise condition, the envelope was low-pass filtered at 400 Hz; this was expected to provide a relatively weak indication of pitch and periodicity.

The vocoder using a pulse source that followed the input fundamental frequency gave substantially higher performance in identification of frequency glides than vocoders using noise carriers, which in turn showed better performance than processors using a fixed rate pulse carrier. However, performance in consonant and vowel identification and sentence recognition was remarkably similar through all of the processors. Connected discourse tracking rates were affected by the envelope filter of the noise carrier processors, although this effect was small. We conclude that whilst the processors achieved the desired control over the salience of pitch and periodicity, the speech tasks used here show little sensitivity to this manipulation.

1. Introduction

Pitch, periodicity and aperiodicity are widely held to be important cues for the perception of speech. However, surprisingly little is known of their exact role in determining speech intelligibility beyond that of pitch for prosody except in what may be a special case, that of auditory signals that contain no spectral structure. With such signals, these factors contribute in several important ways. The timing of periodic and aperiodic excitation are dominant temporal cues to consonant identity (Faulkner & Rosen, 1999). Furthermore, for the audio-visual perception of connected speech, both the timing of voiced excitation and voice pitch provide distinct elements of complementary support to visual cues (Breeuwer & Plomp, 1986; Grant, Ardell, Kuhl, & Sparks, 1985; Rosen, Fourcin, & Moore, 1981). Speech presented through a cochlear implant, or through a vocoder-like simulation of an implant speech processor, is represented by a relatively small number of spectral bands, each

conveying temporal envelope information. It may be expected, then, that the temporal information that contributes to speech perception through such processing is similar to the temporal information that dominates perception from signals that convey no spectral information. This study was performed to discover the effects on speech perception of simulated cochlear implant processors that varied in the salience of the pitch and periodicity information that they conveyed.

Vocoder-like speech processing methods have been used in a number of recent studies that aim to simulate cochlear implant speech processors (Dorman, Loizou, & Rainey, 1997a; 1997b; Rosen, Faulkner, & Wilkinson, 1999; Shannon *et al*., 1995; Shannon, Zeng, & Wygonski, 1998). These simulations represent the spectro-temporal information delivered to the auditory nerve by Continuous Interleaved Sampling (CIS) processors (Wilson *et al*., 1991). In a CIS implant, the signals presented along the electrode array represent amplitude envelopes extracted from a series of band-pass filters. These envelopes, typically smoothed to carry temporal information below 400 Hz, are imposed on bi-phasic pulse carriers that generally have a rate between 1 and 2 kHz.

The simulation studies performed so far have paid little attention to the nature of the temporal cues provided. Rather the focus has been on the role of spectral resolution (Dorman *et al*., 1997b; Shannon *et al*., 1995) and the effects of shifts of the spectral envelope (Dorman *et al*., 1997a; Rosen *et al*., 1999; Shannon *et al*., 1998). Most commonly, simulation studies have made use of band-pass filtered noise carriers to deliver amplitude envelope information to the normal ear. The frequency content of the noise controls the cochlear location to which the information is presented.

Temporal cues to pitch, and to the presence of periodicity, are carried by the modulation of the pulse stimulation from a CIS processor as long as the envelope smoothing filter extends high enough to encompass the voice fundamental frequency range. Similarly, where simulations using vocoder processing use sufficiently high envelope bandwidths to modulate noise carriers, these too are capable of signalling pitch and periodicity for modulation rates up to a few hundred Hz (e.g., Pollack, 1969). However, the salience of the pitch of modulated noise is weak compared to that of harmonic sounds such as voiced speech, and it is important to establish the limitations that such simulations may have in respect of the transmission of pitch and periodicity information. Little is presently known about the effects of the salience of periodicity in such simulations. Fu & Shannon (2000) report little effect of varying the envelope cut-off frequency between 16 and 400 Hz for English consonant materials with four-channel noise-excited vocoders. In Chinese, however, it has been shown that tonal cues carried by noise modulated by a 400 Hz bandwidth speech envelope can contribute to sentence level speech perception using such simulations (Fu, Zeng, & Shannon, 1998).

There are two related issues addressed here. Firstly, given that simulations of vocoder-like CIS speech processors deliver limited pitch and periodicity information, what impact does this have on standard measures of speech intelligibility, and would more salient pitch and periodicity cues improve performance? Closely linked to this first issue is the question of the adequacy of current methods of simulating CIS cochlear implant speech processors in relation to the ability of implant patients to derive pitch and periodicity information from CIS processors.

1.1 Pitch and periodicity cues from a CIS cochlear implant processor

The representation of pitch and periodicity for users of a CIS cochlear implant speech processor will depend both on the extent to which the corresponding temporal information is contained in the extracted amplitude envelopes, and on the extent to which the patient is able to process this information. This latter aspect is not well understood, although it is clear that there are very wide variations between patients. A study of periodic/aperiodic discrimination in single channel implant users showed some patients to have good abilities in identifying periodic from aperiodic pulse stimulation at least for stimuli of 200 ms duration (Fourcin et al., 1979). However, except for one subject, the stimuli used in that study were directly periodic or aperiodic, not pulse carriers with periodic or aperiodic amplitude modulation. McDermott and McKay (1997) studied one individual implant patient under conditions comparable to CIS stimulation. Sinusoidal amplitude modulation of a 1200 Hz pulse train delivered to a single bipolar electrode pair allowed the discrimination of modulation rates differing by 3% to 4% around a 100 Hz rate. Around a 200 Hz rate, thresholds were between 4 and 27%, depending on the stimulation site. Other selected CIS implant processor users have also showed good ability in the pitch ranking of pulsatile stimulation that carries sinusoidal amplitude modulation (SAM) up to modulation rates of up to 1 kHz (Wilson *et al*., 1997). However, this study gives rather limited information on pitch discrimination, since the ranked modulation rates differed in steps of 100 Hz.

1.2 Representation of pitch information in vocoder carriers

In normal hearing, pitch perception is thought to be based both on cues derived from resolved lower frequency harmonics, including the fundamental component, and also from periodicity cues in the temporal envelope in auditory filter channels driven by adjacent unresolved harmonics. A CIS implant processor cannot deliver place pitch cues within the voice fundamental frequency (F_x) range, nor preserve pitch cues from individual harmonics of a complex signal, since the channel band-pass filters are too wide. Hence only envelope periodicity cues will be available. The carrier in a CIS processor is a non-random high rate pulse rather than the random noise typically used in simulations. For this reason, temporal modulation of the carrier related to F_x will be noise-free, and the neural responses to this stimulation are also likely to be strongly synchronised to the modulation (Wilson *et al*., 1997).

This study introduces the use of frequency-controlled pulse carriers for voiced speech. Here the carrier for voiced speech is a flat spectrum pulse train whose period is controlled by voice fundamental frequency. The carrier is passed through a series of band pass filters to determine the frequency content of the different output bands. The use of such a carrier is not intended to represent the pulsatile stimulation of CIS, which indeed cannot be accurately emulated in acoustic hearing. Rather the intention is to achieve the highest possible pitch salience by providing a rich set of pitch cues both from individually resolved lower harmonics and from temporal envelope cues from the unresolvable higher harmonics. The noise carriers typical of most previous simulation studies necessarily lack harmonic content, and provide only temporal envelope cues to pitch. Here the periodicity of the temporal envelope related to voice pitch will be noisy by virtue of the random nature of the carrier. These random fluctuations in the carrier level will be more significant in the lower vocoder bands, where the rate of the inherent envelope fluctuations of the carrier is closer to the rate

of the envelope fluctuations extracted from periodic speech. Such effects will be still greater if the carrier bandwidth is within the voice fundamental frequency range. Because of the random nature of such noise carriers, the resulting pitch salience may be weaker than for that derived from CIS processors by implant users who are able to fully process the temporal information carried by envelope-modulated pulse stimulation.

2. Methods

2.1 Signal processing

Signal processing was implemented in real-time, using the Aladdin Interactive DSP Workbench software (v1.02, AB Nyvalla DSP). It ran at a 16 kHz sample rate on a Loughborough Sound Images DSP card with a Texas Instruments TMS320C31 processor. All processors used here had four channels, with the analysis and output filters being identical, so that the spectral representation was tonotopically accurate within the constraints of the limited spectral resolution. A block diagram of the common components of the processors is shown in Figure 1. Each channel consisted of a series of blocks, comprising; a band-pass filter applied to the speech input, a rectifier and low-pass filter to extract the amplitude envelope from that spectral band, a multiplier that modulated a carrier signal by that envelope, and a second band-pass filter, matching the analysis filter, to shape the spectrum of the modulated carrier signal.

The four analysis and output filter bands were based on equal basilar membrane distance, with the lowest frequency band having its lower cut-off frequency at 100 Hz. Filter cut-off frequencies at the –6dB point are shown in Figure 2. These were calculated from the formula of Greenwood (1990). The band-pass analysis filters, and the corresponding output filters, were 8th order elliptical IIR designs, with slopes in excess of 50 dB/octave, and stop-bands at least 50 dB down on the pass-band.

The amplitude envelope was extracted from each analysis filter output by half-wave rectification followed by a 4th order elliptical low-pass filter, with a slope of about 48 dB/octave. The analysis and envelope filters were designed using QEDesign1000 v3.04 for Windows.

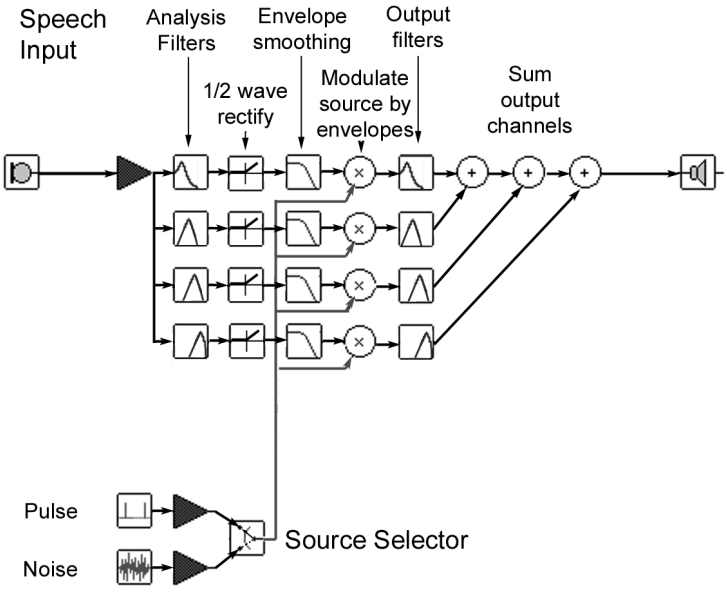

Figure 1. Block diagram showing speech processing common to all processor simulations.

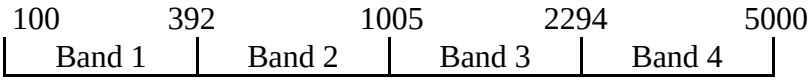

Figure 2: Channel cut-off frequencies (Hz)

2.1.1 Speech processing conditions

The various processing conditions are summarised in Table I. The fullest and most salient representation of pitch and periodicity was produced using processing similar to classic speech synthesising vocoders (Dudley, 1939). Here the carrier source during voiced speech was a pulse signal whose frequency followed that of the fundamental frequency of the speech input, $F_{x.}$ The carrier source for voiceless speech was a random noise (symbolised as Nx). This condition is notated as **FxNx**. The pulse carrier was a mono-phasic pulse with a width of one sample (63 μS). Within the 8 kHz overall bandwidth of the processor, the spectral envelope of this pulse train and the noise source were both flat, and both source signals had the same *rms* level.

Processor	Voiced speech carrier	Voiceless speech carrier	Envelope low-pass cut-off (Hz)	Expected salience of pitch and periodicity
Noise400	Noise	Noise	400	Both weak
Noise32	Noise	Noise	32	Nil
VxNx	150 Hz pulse train	Noise	32	High for periodicity, nil for pitch
FxNx	F_x pulse train	Noise	32	Both high
Mpulses	150 Hz pulse train	150 Hz pulse train	32	Nil

Table I: Summary of processor conditions (see text for details)

A processor similar to that used for condition **FxNx** differed only in using a fixed 150 Hz pulse rate rather than a speech-derived pulse rate. This processor preserved the contrast between periodic and purely aperiodic excitation, while discarding voice pitch. It is designated as condition **VxNx**.

A third processor discarded both periodicity and pitch information and was produced by using a fixed frequency 150 Hz pulse source for all speech input. This condition was designated **Mpulses** (monotone pulses).

Two further processors employed a filtered white noise carrier for all speech. These are similar to the processors used by Shannon *et al.* (1995). They differed from each other only in the low-pass cut-off frequency of the envelope filters, which was either 400 Hz in condition **Noise400**, or 32 Hz in condition **Noise32**. A 400 Hz envelope cut-off is typical of commercial CIS speech processors, and was expected to allow periodicity and pitch to be preserved in the extracted envelope. However, the perceptual salience of this information was not expected to be as high as for condition **FxNx**. The use of a 32 Hz cut-off frequency together with the 48 dB/octave slope of the envelope filter was expected to eliminate virtually all pitch and periodicity cues in condition **Noise32**. The higher envelope cut-off of processor **Noise400** also allows more rapid between-channel spectral changes to be represented compared to all the

other processors. The processors **Noise32** and **Mpulses** represent virtually identical spectro-temporal information, and differ only in that the output is either always aperiodic or always periodic.

2.1.2 Voicing detection and source switching

All speech materials were accompanied by a laryngographic signal marking glottal closure. Before processing through the simulations, the raw laryngograph waveform was pre-processed to produce a single discrete pulse at each laryngeal closure. The processors took this pulse train as input in addition to the speech signal. A DC offset was added to the pulse input to ensure that it passed through zero, and a zero-crossing detector was employed to detect the pulse period. Alternate zero-crossings triggered the generation of a carrier pulse. A sample and hold with a 10 ms time constant was applied to the output of the zero-crossing detector and the output of this stage was used as a voicing detector. The voicing detector output, smoothed by a 1st order 50 Hz low-pass filter, was used to switch between the pulse and a white noise source. The input speech was delayed before the initial band-pass analysis filtering by 30 ms to allow accurate time alignment of the switching between the vocoder carrier signals with changing speech excitation.

2.2 Results of speech processing

Figure 3 shows the output of processor **FxNx** for six alveolar intervocalic consonants in an /ɑ/ vowel context together with the original speech. Formant structure and other spectral details are absent on the processed signal, but F_x, and the timing of periodicity and aperiodicity, are well preserved. Note that during the voiced fricative /z/, and at the release of the voiced stop /d/, the presence of laryngeal excitation always results in the selection of the pulsatile source, even when aperiodic excitation is also present in the speech signal.

2.3 Speech perceptual tests

Speech performance for segmental and connected-speech materials was measured using four standard procedures. All speech tests used auditory presentation.

2.3.1 Consonant identification

The consonant set contained 20 intervocalic consonants with the vowel /A/. These comprised all the English consonants except for /ð, ʒ, h, ŋ/. Materials were from digital anechoic recordings presented at a 22.05 kHz sample rate and were from one female and one male talker, mixed in each test run. Both talkers had a standard Southern British English accent. Each run presented 40 consonants selected at random from a total set of six to ten of each token from each talker. Stimulus presentation was computer-controlled. Subjects responded using the computer mouse to select one of 20 buttons on the computer screen that were orthographically labelled to represent each of the twenty consonants.

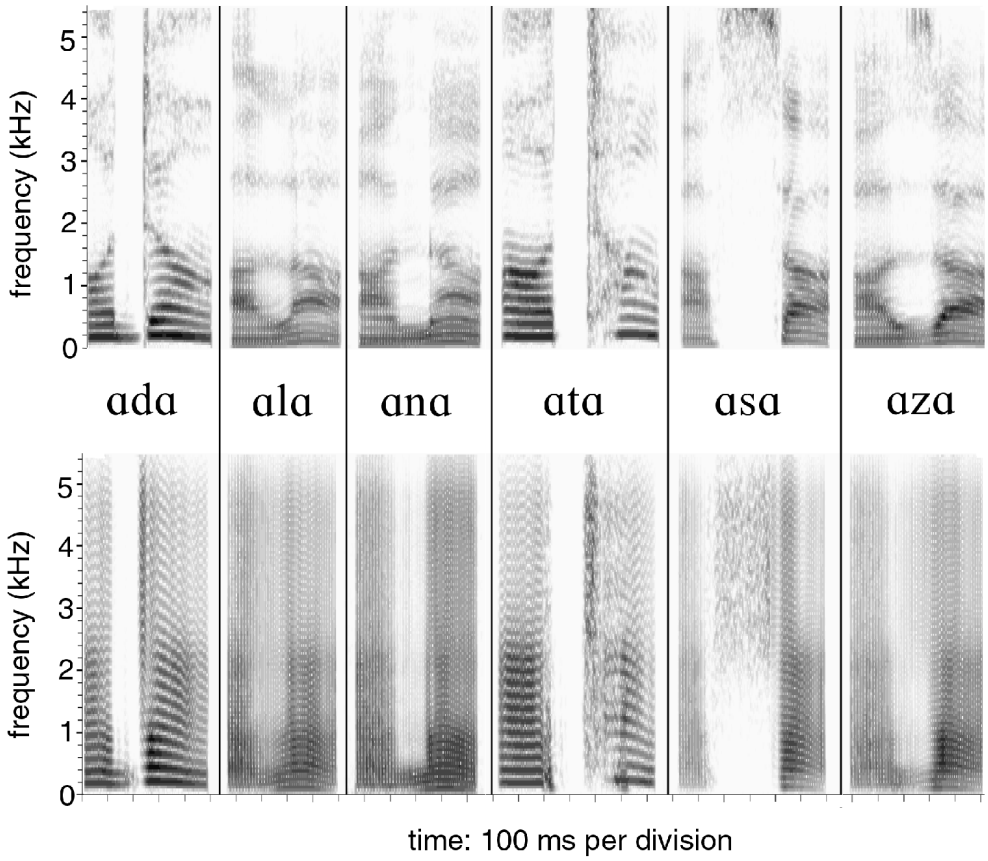

Figure 3: Spectrograms of speech input (upper panels) and output of processor FxNx (lower panels) for six alveolar consonants. The spectrograms use a 100 Hz bandwidth. The /ada/ and /ata/ tokens are from the male talker, the remainder from the female talker. Each panel shows the transition regions from and into the vowel.

2.3.2 Vowel identification

17 b-vowel-d words from the same two talkers were used, again from digital anechoic recordings presented at a 22.05 kHz sample rate. Presentation was computer-controlled. Each test run presented one token of each word from each of the two talkers, selected at random from a total set of six to ten of each token from each talker. The vowel set contained ten monophthongs (in the words bad, bard, bead, bed, bid, bird, bod, board, booed, and bud) and seven diphthongs (in the words bared, bayed, beard, bide, bode, boughed, and Boyd). The spellings given here were those that appeared on the computer response buttons.

2.3.3 Sentence perception

BKB sentences from a different female talker with the same British accent were used, from an analogue audio-visual recording on U-matic videotape (EPI Group, 1986; Foster *et al*., 1993). Each test run used one list of 16 sentences with 50 scored key words per list.

2.3.4 Connected Discourse Tracking

Live voice connected discourse tracking (CDT: DeFilippo & Scott, 1978) was conducted by a third single female talker (author CS). In CDT, the talker wore laryngograph electrodes to provide a larynx period and voicing reference. Materials were taken from texts for students of English as a second language.

2.4 Pitch salience test

In addition to the three speech assessments, pitch salience through each processor was examined by the use of tone glides. The stimuli were sawtooth waves having a spectrum similar to that of voiced speech. Each was 500 ms in duration and had a linear fundamental frequency transition from start-to-end. Three fundamental frequency ranges were included, centred around 155, 220, and 310 Hz. The start and end frequencies of the glides varied in 6 steps from a ratio of 1:0.5 to a ratio of 1:0.93. The test was again presented under computer control. On each trial, subjects heard a single glide, and were asked to categorise it as either "rising" or "falling" in pitch. They responded by clicking on one of two response buttons labelled with a rising or falling line. Each single administration of this test presented one rising and one falling tone at each start-to-end frequency ratio in each of the three frequency ranges, 36 stimuli in total.

3. Procedure

Five subjects, screened for normal hearing up to 4 kHz, were recruited for the consonant, vowel, sentence and tone glide tests. For each of these tests, six testing blocks were presented, in which each of the four tests was administered once through each of the five processors. The first two blocks were treated as practice. Because only 21 BKB lists were available, one identical BKB list was presented repeatedly for the first two practice blocks. In the final four blocks, a different BKB list was presented on every occasion.

CDT was run subsequently, with two subjects who had taken part in the earlier tests and an additional four subjects who were also screened for normal hearing. The CDT testing used only four of the five processors, with the **VxNx** processor being excluded. Each of the total of six testing sessions included 10 minutes of CDT with each of the four processors used. Each of these 10 minute blocks was scored in two sub-units of five minutes duration. The order of use of the four processors was counterbalanced in a different order for each subject over the six test sessions.

4. Results

4.1 Frequency glides

Psychometric functions for labelling of glide direction as a function of start-to-end frequency ratio are shown in Figure 4. For processor **FxNx**, performance for all the glide stimuli was at very high levels. Even with the smallest start-to-end frequency ratio of 1:0.93, scores were around 90% correct. Both modulated-noise processors allowed a limited identification of the direction of pitch glides. Performance with processor **Noise400** was above 75% correct for ratios of 1:0.66 and larger. Performance with processor **Noise32** was poorer than with **Noise400**, but better than that shown by the fixed frequency pulse processors. For these, performance was close

to chance as would be expected. The above chance performance with processors **VxNx** and **Mpulses** at the largest frequency ratios can be attributed to spectral envelope differences that arise as harmonics of the input signal shift between processor bands.

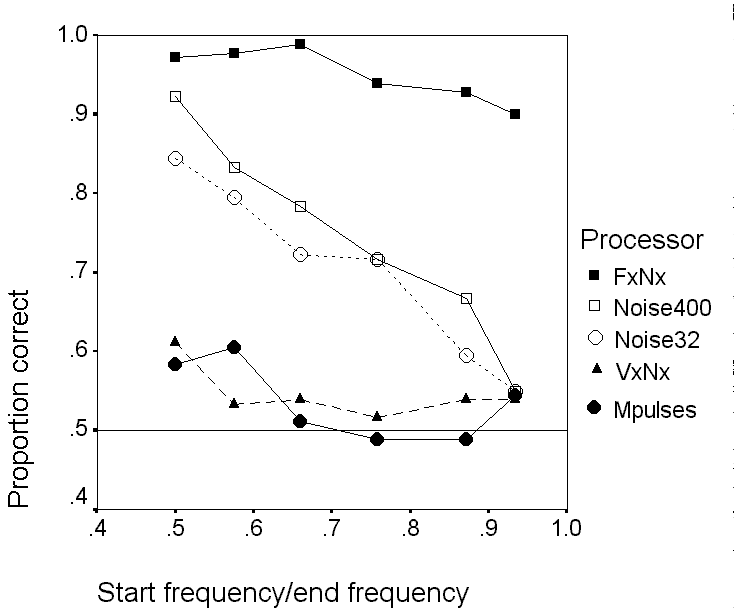

Figure 4: Performance in labelling processed sawtooth wave frequency glides as a function of the ratio of the start and end frequencies (ignoring glide direction). Each point shows a mean score over four test sessions from 15 samples over the 3 frequency ranges and the 5 subjects. Chance performance is 50% correct.

Psychometric functions for the proportion of "fall" responses as a function of the log(base10) of the start to end frequency ratio were estimated using a logistic regression applied to the group data. The resulting slope estimates and their 95% confidence limits are shown in Figure 5. The slope for processor **FxNx** is substantially steeper than that in all other conditions. The slope for the 400 Hz envelope bandwidth noise carrier processor **Noise400** is slightly but significantly steeper than that for the **Noise32** processor. Slopes for the two fixed period pulse processors **Mpulses** and **VxNx** are close to zero.

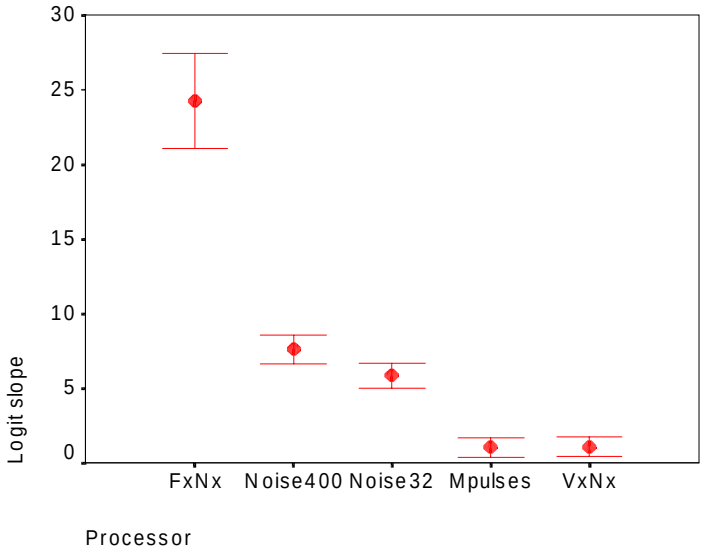

Figure 5. Slopes of the psychometric functions estimated from a logistic regression of proportion "Fall" responses as a function of the log(base10) of the start to end frequency ratio. Error bars are 95% confidence limits.

4.1.1 Contribution of temporal envelope and spectral cues to glide labelling

A more detailed investigation of the pitch cues available from the noise-carrier processors has been carried out that also permits a dissociation of the contributions of temporal envelope and spectral cues. Here, in addition to processors **Noise32** and **Noise400**, two single-band processors with 32 Hz and 400 Hz envelope bandwidths were employed. These were identical to **Noise32** and **Noise400**, except that the four analysis filters and envelope extractors were replaced by a 50 Hz high pass filter followed by a single envelope extractor whose output modulated the level of each of the four output bands. With these single wide-band analysing processors, the output always had a constant level over the four bands. Hence the output was unaffected by the number of harmonics of the sawtooth signal falling in each analysis band, and spectral cues correlated with F_x were eliminated. A single subject (author AF) took part in this experiment. A total of 13 test sessions with each of the four processors were run, with order randomised between processors. Data were fit by logistic regression of the proportion of "rise" responses against the log (base 10) of the ratio of the start-to-end frequency of the stimuli (13 observations per point). Fits were performed for each of the three frequency ranges (centred around 155, 220 and 310 Hz), and for the four processors. None of the fits deviated significantly from the observed data according to chi-squared. The slopes of the logit fits are shown in Figure 6. The intercepts of the fits were all similar and were slightly less than zero, indicating an overall bias towards the response "fall".

For the single wide-band processor and the 32 Hz envelope bandwidth (left panel of Figure 6), the slopes were very close to zero for all three glide centre frequency ranges. For these processors there is around a 0.2 change in the probability of the response "rise" between a rise and a fall of one octave. There is no spectral information available from this processor, nor any differential temporal envelope information across the three glide frequency ranges. For the 400 Hz envelope bandwidth single-band processor, the slopes of the psychometric functions showed an orderly and strong relation to the glide frequency range. For glides centred on 155 Hz, the slope was fairly high, representing a change of approximately 0.8 in the probability of the response "rise" as the glide start-to-end frequency ratio changes between 0.9 and 1.1. For glides centred on 310 Hz, however, the slope is as low as for the 32 Hz envelope bandwidth processors. This relation between glide centre frequency and the slope of the psychometric function reflects the declining utility of envelope cues to pitch between 155 and 310 Hz.

The data in the right panel of Figure 6 show the combined effects of spectral envelope and temporal envelope cues from the processors. With a 32 Hz envelope bandwidth, once more there was no effect of glide frequency range, but the slopes are clearly steeper than those from the single wide-band processor. Here there is approximately an 0.55 change in the response probability as the glide start-to-end frequency changes from 0.9 to 1.1. Essentially the same slopes were seen for a 400 Hz envelope bandwidth as for the 32 Hz bandwidth for glides centred around 220 and 310 Hz. This level of performance can be attributed purely to the spectral differences that arise as the harmonics of the glide stimuli move between processor bands. These slopes are also very similar to that for the glides around 220 Hz for the single band 400 Hz envelope bandwidth condition. This suggests that there is little integration of cues from spectral and temporal envelope sources. Finally, the glides centred around 155 Hz yielded a very steep psychometric function for the four band processor when the

envelope bandwidth is 400 Hz. This slope exceeds that for the same glide frequency range and envelope bandwidth for the single-band processor, although not significantly.

In conclusion, both spectral and temporal cues are contributing to glide labelling performance with the four-band 400 Hz envelope bandwidth processor **Noise400**. There is a clear effect of fundamental frequency range on the salience of envelope cues to the pitch of amplitude modulated noise. Around a 155 Hz centre frequency, envelope cues are strong, but as the centre frequency reaches 310 Hz, the envelope cues contribute virtually nothing to pitch glide labelling performance. Spectral cues, but not temporal cues, contribute to above-chance pitch glide labelling with the four-band 32 Hz envelope bandwidth processor **Noise32**. Where a series of carrier bands are modulated by envelopes extracted from speech, rather than from a steady-state periodic tone, such spectral cues to pitch are likely to be obscured by the time-varying spectral envelope of speech. As a result, spectral pitch cues would not be likely to be useful when speech is processed in this way.

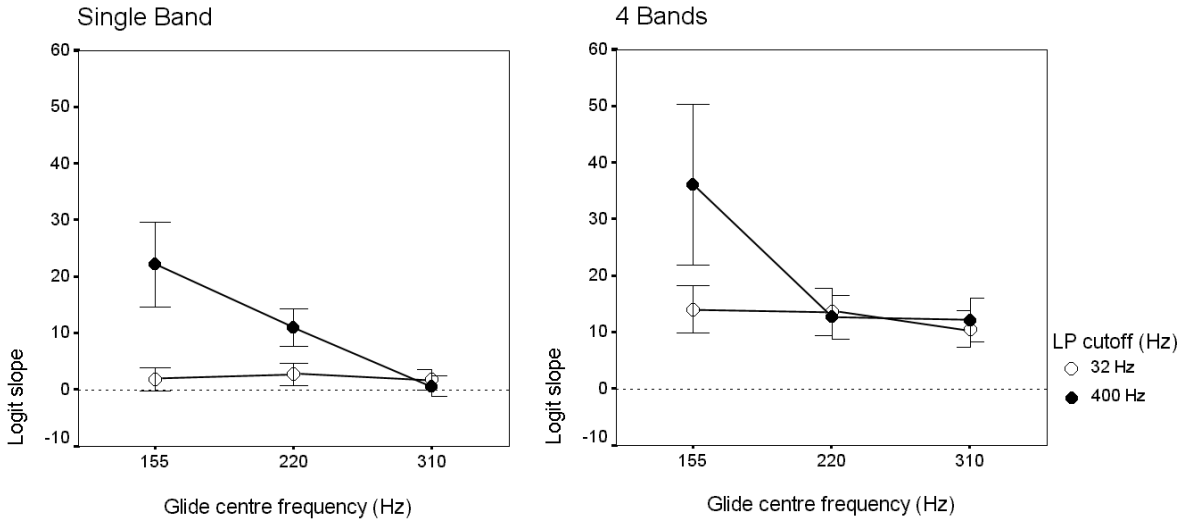

Figure 6: Slopes of logistic regression fits to psychometric functions of glide labelling against the log of the start-to-end frequency ratio. Increasingly positive slopes represent steeper psychometric functions and higher performance. Error bars show 95% confidence limits for the logistic regression slope estimate.

4.2 Comparison of pitch labelling data with patient performance

Measures of pitch glide labelling have also been made in six patients using the MEDEL CIS-link processor with the Ineraid electrode array. These measures used the same stimuli and a similar presentation method to the simulation results presented above. Only two of these patients could confidently label the direction of pitch glides imposed on saw-tooth waves. The remaining four were unable to identify the direction of one octave glides centred around 155 Hz, and could also not discriminate rising and falling one octave glides in a same-different task. For those two patients who were able label tone glides, psychometric functions are shown in Figure 7 together with functions from the **FxNx** and **Noise400** simulations.

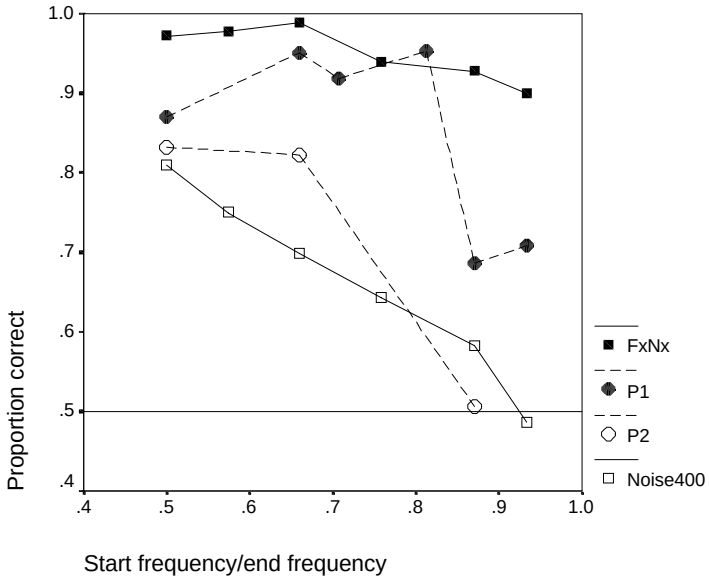

Figure 7, Psychometric functions for labelling pitch glide direction. Simulation results are shown as squares, data from patients as circles.

Patient P2 showed performance comparable to that from processor **Noise400**. However, patient P1 performed substantially better than simulation results from this processor, at levels approaching those from the **FxNx** simulation. That this level of accuracy in pitch glide labelling can be achieved, if only by one of this sample of six patients, indicates that noise carrier-based vocoders do underestimate the salience of pitch information to some users of CIS speech processors.

4.3 Vowel identification

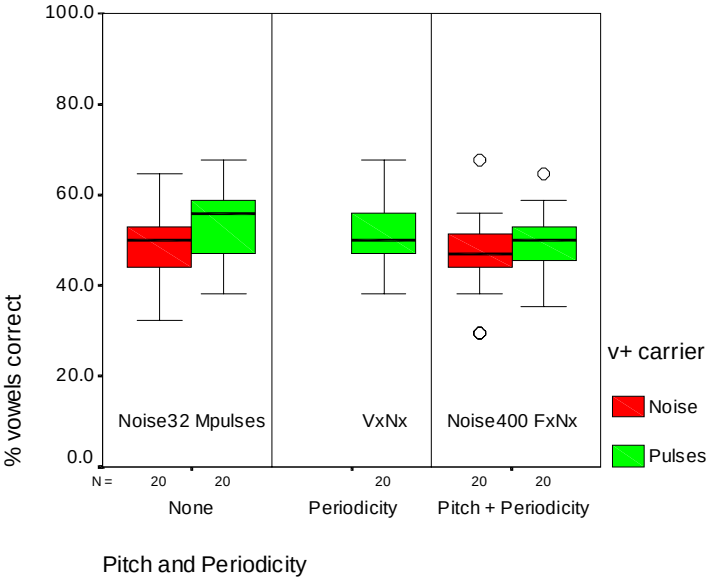

Figure 8: Percent correct vowel identification in the processor conditions. Boxes represent the 25% to 75% ranges of the data (over subjects, talker, and test run). The bar within each box is the median. The whiskers show the range of scores excluding any points that are more the three box widths from the median; these are shown as open circles

Box and whisker plots of the group data with each processor are shown in Figure 8. Scores were around 50% correct in all conditions. A repeated measures ANOVA was carried out on data from the last four test sessions using a factorial structuring of the processor conditions (omitting the **VxNx** processor). One factor represented the presence or absence of pitch and periodicity and contrasted processors **FxNx** and

Noise400 with **Mpulses** and **Noise32**. A second factor represented the use of a pulse or a noise carrier for voiced speech, and contrasted processors **FxNx** and **Mpulses** with **Noise32** and **Noise400**. Talker and test session were additional factors. The only significant effect was that of test session [F(1,3) = 13.68, p = 0.0004, power 0.998][i]. That processors conveying voice pitch did not lead to higher scores is perhaps surprising. The male and female speakers differ in formant frequency range, and processors that signal voice fundamental frequency and hence speaker sex might be expected to show higher performance. Performance with these four channel processors is comparable to that found for a processor similar to **Noise400** in a previous study using the same vowel set from the female talker only (Rosen, Faulkner, & Wilkinson, 1997, 1999). A second ANOVA that treated the processors as 5 levels of a single factor included the **VxNx** condition. This showed no significant processor effect.

4.4 Intervocalic consonants

4.4.1 Overall accuracy

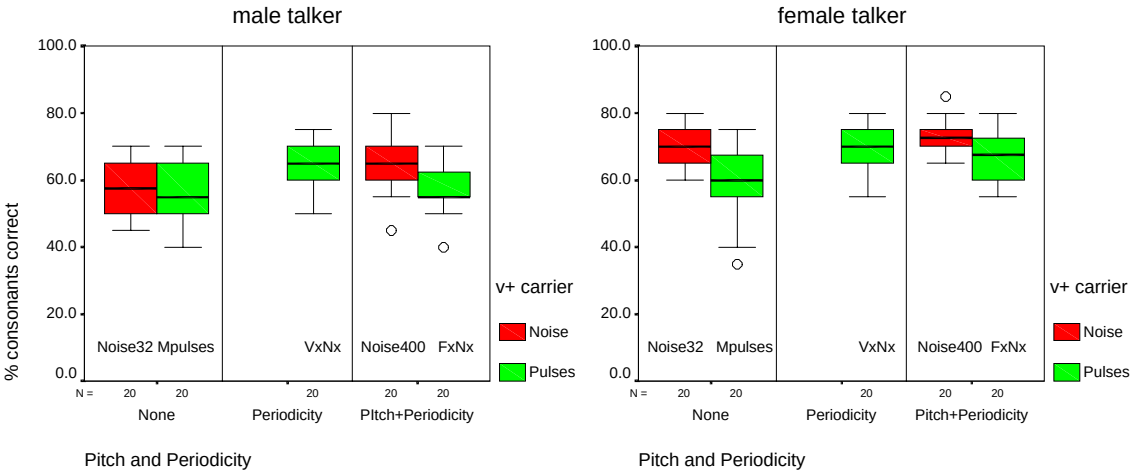

Figure 9: percentage correct consonant identification for each talker using the five processors. Box plots show the distribution of scores over subject and test run.

Group results are shown in Figure 9. A repeated measures ANOVA of overall accuracy was carried out again using a factorial partition of the processor conditions excluding **VxNx**. Factors were the presence or absence of periodicity and pitch cues, the carrier used for voiced speech, talker and test session. Here there was no effect of test session, nor any interactions with this factor. Talker was a significant main effect [F(1,4) = 50.45, p = 0.002, power = 1.000]. The presence of periodicity and pitch information was also a significant factor [F(1,4) = 159.4, p =0.0002, power = 1.000], indicating that average scores through processors **FxNx** and **Noise400** significantly exceeded average scores through processors **Mpulses** and **Noise32**. There was no significant effect of the carrier factor. There was, however a significant higher-order interaction between the periodicity, carrier and talker factors [F(1,4) = 8.409, p = 0.044, power = 0.592]. This interaction suggests a larger effect of periodicity for the noise carrier for the male talker than for the female talker, and conversely, a larger effect of periodicity for a pulse carrier for the female than for the male talker.

Bonferroni paired comparisons from a similar ANOVA in which all processor conditions were treated as levels of a single factor were carried out to look for differences between individual conditions. These showed no significant pairwise differences between conditions.

4.4.2 Consonant Feature Information

A second series of ANOVAs were performed on information transfer measures (Miller & Nicely, 1955) computed from confusion matrices summed over the last four test sessions. These data are displayed in Figure 10. A summary of these ANOVAs is presented in Table II.

Measure	*Factor*	F	p	Eta ^{2}	Observed Power
Voicing	*Periodicity*	23.06	**0.0086**	0.852	0.941
	Carrier	12.79	**0.0232**	0.762	0.762
	Talker	7.60	0.051	0.655	0.551
	*Periodicity*Carrier*	16.15	**0.0159**	0.801	0.846
	*Periodicity*Talker*	3.93	0.118	0.496	0.332
	*Carrier*Talker*	3.17	0.149	0.442	0.279
	*Periodicity*Carrier*Talker*	1.053	0.363	0.209	0.126
Place	*Periodicity*	0.957	0.384	0.193	0.119
	Carrier	16.7	**0.015**	0.807	0.857
	Talker	32.6	**0.0047**	0.891	0.985
	*Periodicity*Carrier*	11.9	**0.0259**	0.749	0.735
	*Periodicity*Talker*	0.140	0.727	0.034	0.060
	*Carrier*Talker*	0.191	0.684	0.046	0.064
	*Periodicity*Carrier*Talker*	3.80	0.123	0.487	0.322
Manner	*Periodicity*	1.84	0.247	0.315	0.183
	Carrier	5.67	0.0759	0.586	0.444
	Talker	13.7	**0.0209**	0.774	0.787
	*Periodicity*Carrier*	0.046	0.841	0.011	0.053
	*Periodicity*Talker*	0.541	0.503	0.119	0.089
	*Carrier*Talker*	0.126	0.740	0.032	0.059
	*Periodicity*Carrier*Talker*	0.217	0.666	0.051	0.065

Table II: Summary of Repeated Measures ANOVAs of consonant feature information transmission. All F tests had 1 and 4 degrees of freedom.

A more salient representation of periodic and aperiodic excitation would be expected to lead to improved identification of manner and voicing features (Faulkner, Potter, Ball, & Rosen, 1989; Faulkner & Rosen, 1999). An ANOVA of voicing information scores did indeed show a significant effect of the representation of pitch and periodicity. There was also a significant effect of the carrier for voiced speech, and an interaction between the carrier and the representation of periodicity. Voicing scores through the processors conveying pitch and periodicity (**FxNx** and **Noise400**) were almost all at 100%. Scores using processors **Noise32** and **VxNx** were somewhat lower, while those from processor **Mpulses** were substantially lower than through the other processors. Bonferroni-corrected comparisons from an ANOVA using a single

processor factor showed that voicing scores through processor **Mpulses** were significantly lower than through all other processors. No other pair-wise differences reached significance.

For manner information there were no significant effects of processor, only an effect of talker. This suggests that the periodic/aperiodic contrast is not a powerful cue for signalling feature contrasts such as that between voiceless fricatives and voiced plosives or nasals, despite the difference in the excitation sources.

There were significant main effects of talker and carrier for place information, and as for voicing information, an interaction of the periodicity and carrier factors (see Table II). Bonferroni-corrected comparisons showed the processors **FxNx** and **Mpulses**, for which mean place scores were lowest, to be significantly different from processor **Noise400**, which led to the highest place information. While all processors except for **Noise400** presented equivalent spectro-temporal information, **Noise400** represented more rapid spectral envelope changes (at rates between 32 and 400 Hz) which were not present in the output of the other processors. This seems the most likely explanation for the higher place scores obtained through this processor.

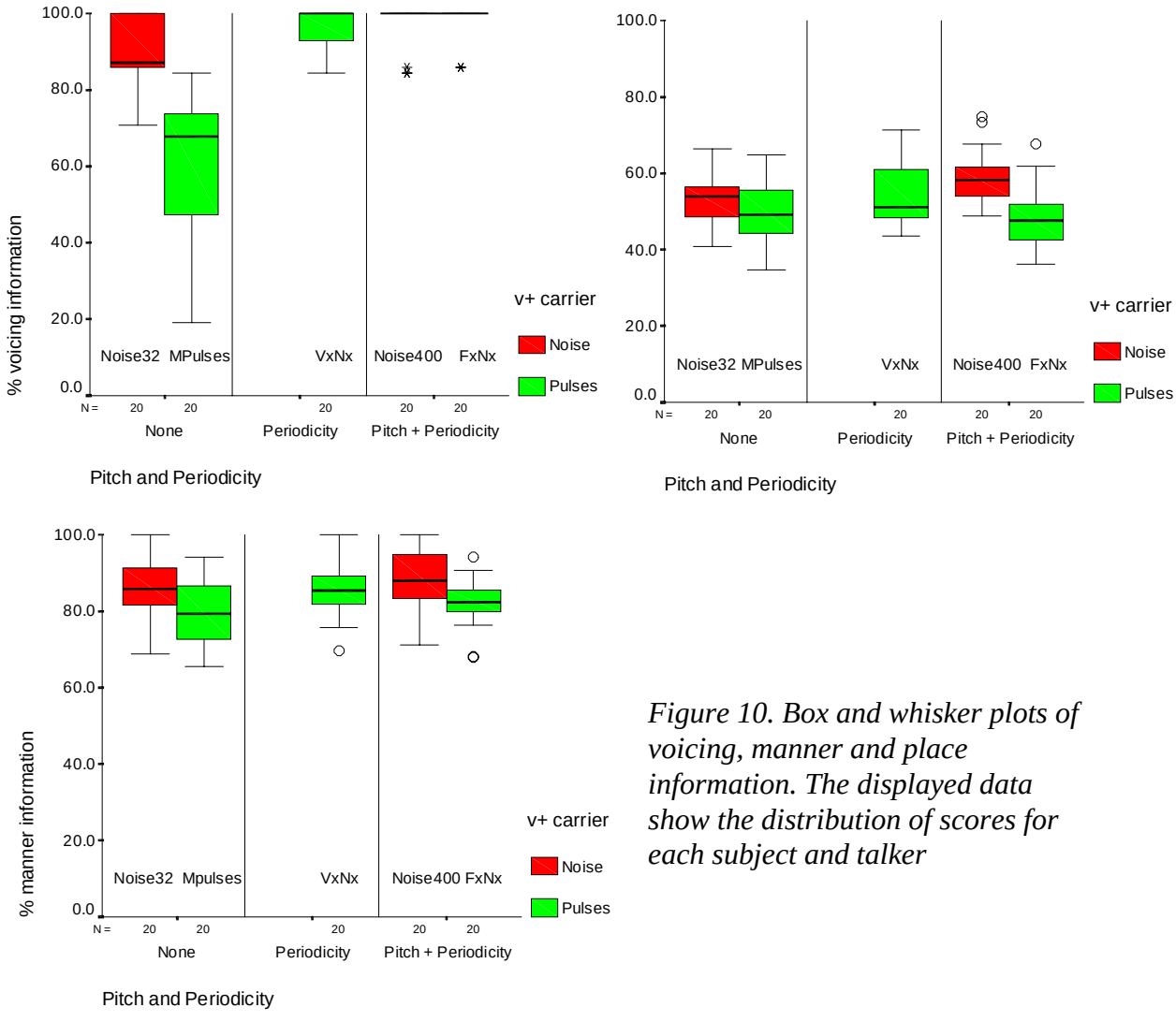

Figure 10. Box and whisker plots of voicing, manner and place information. The displayed data show the distribution of scores for each subject and talker

4.5 BKB sentences

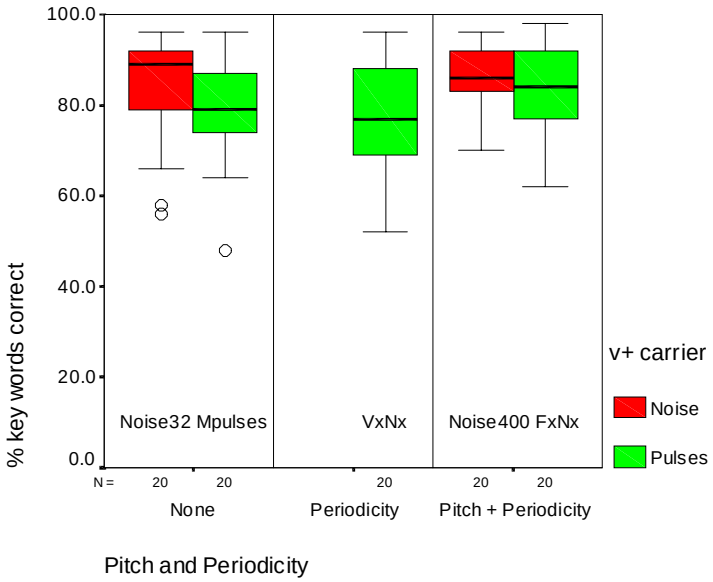

Figure 11. Percentage of key words correctly identified from BKB sentences

Group scores using the key-word loose scoring method are shown in Figure 11. Scores were high for a four channel processor compared to another study that used the same materials and a processor similar to the **Noise400** condition (Rosen *et al*., 1997, 1999), and may be limited by ceiling effects. A repeated measures ANOVA using factors of the representation of pitch and periodicity, carrier for voiced speech, and test session was performed. This showed only a significant main effect of carrier type [$F(1,4) = 40.4$, $p = 0.003$, power = 0.995], indicating that scores through the noise carrier processors (**Noise32** and **Noise400**) were higher than those through processors using a pulse carrier for voiced speech (**FxNx** and **Mpulses**). Bonferonni paired comparisons between all five processors from an ANOVA using a single processor factor showed only one pairwise difference, this being between the highest scoring processor **Noise400** and the lowest, **VxNx**.

4.6 Connected Discourse Tracking

The **VxNx** processor was not used for CDT. Unprocessed speech was used in the initial CDT sessions with each subject both to familiarise them with the task and to estimate ceiling performance rates. Tracking rates through the other four processors were all significantly lower than that with unprocessed speech. A repeated measures ANOVA was applied to CDT rates over the last four 10 minute testing blocks with each processor, excluding the unprocessed speech condition. This showed a significant effect of block [$F(1.96, 9.78) = 8.22$, $p = 0.008$, power = 0.875]. Block did not interact with any other factor. There were no main effects of periodicity or carrier, but there was a significant periodicity*carrier interaction [$F(1,4) = 7.72$, $p = 0.039$, power = 0.608]. Bonferroni corrected paired comparisons from a similar ANOVA that used a single processor factor showed that rates through processor **Noise400** were significantly higher than the slowest rates, obtained from processor **Noise32**. The mean difference between these two conditions was 5.95 words/minute, while the standard deviation of this difference was 5.01. No other pair-wise differences between processors were significant.

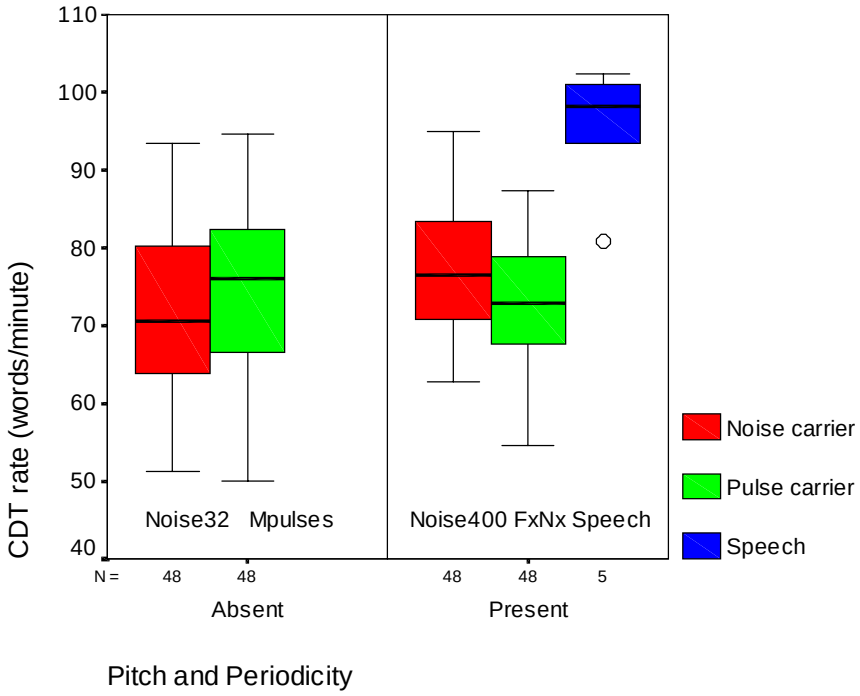

Figure 12: CDT rates for four processors and for unprocessed speech. Data for speech are missing for one subject.

That the noise carrier processors showed a significant effect of the envelope filter cut-off suggests that pitch and periodicity cues may increase the ease and rate of speech communication. However, since rates through processor **FxNx** did not exceed those through processor **Mpulses**, it may be that the difference between the **Noise400** and **Noise32** processors is due to the signalling of more rapid spectral changes by processor **Noise400** rather than to the presence of pitch and periodicity cues.

5. Conclusions

5.1 Role of pitch and periodicity in simulated processors

The consonant, vowel, and sentence tests showed only small effects between processors. That significant effects were found confirms that the design has sufficient statistical power to detect differences. The processor with an Fx controlled pulse carrier rate never led to significantly higher scores than either of the purely noise carrier processors. This was the case even when the noise was modulated by an envelope containing no information above 32 Hz. We may conclude that despite the limited salience of pitch and periodicity, that modulated noise carriers are adequate for the simulation of cochlear implant processors for speech intelligibility tasks such as those used here. Conversely the limited sensitivity of any of the speech measures here to variations of pitch salience may signal the inadequacy of all of these measures for evaluating the availability to a listener of the full range of significant acoustic factors in speech perception.

5.2 Temporal and spectral pitch and periodicity cues

Results from the frequency glide labelling task confirm that processors differ in the salience of pitch information. The supplementary study of frequency glide labelling as a function of spectral and temporal cues shows, as should be expected, that processors

with a 32 Hz envelope bandwidth cannot convey temporal envelope cues to pitch in the voice fundamental frequency range. It seems unlikely that temporal cues to periodicity/aperiodicity would be available when temporal cues to pitch are absent, since periodicity information necessarily resides in the same envelope frequency range as pitch information. Processors with a 32 Hz envelope bandwidth can signal spectral envelope shifts that, with the sawtooth wave stimuli used here, are correlated with fundamental frequency change, and can weakly signal pitch glide direction. It seems unlikely, however, that with a signal such as speech, whose spectrum is constantly changing, there exist spectral shifts that are sufficiently well correlated with fundamental frequency to signal voice pitch change in the absence of more salient temporal cues or of resolved harmonic components.

Processors **Noise32** and **Mpulses** differed only in the use of a noise compared to a pulsatile carrier, and apart from the random nature of the noise carrier, they conveyed identical spectral and temporal information, with temporal cues to pitch here being negligible. However, scores from processor **Mpulses** were significantly lower than from processor **Noise32** for the frequency glide task. This suggests that a carrier with a salient and constant pitch may in some way "mask" pitch cues carried in the spectral information provided by the processor.

Pitch labelling data from cochlear implant patients suggest that at least some patients, such as P1 in the present study, can extract pitch cues from CIS stimulation with a degree of precision that substantially exceeds that possible for the extraction of pitch cues from modulated noise by normally hearing listeners.

5.3 Speech information from periodic and aperiodic carriers

The explicit encoding of periodicity and aperiodicity in the carrier signals affected only the transmission of consonant voicing. Here, processors **FxNx** and **VxNx** led to significantly higher scores than processor **Mpulses**, where the carrier was always periodic. However, the noise-based processors also showed higher transmission of voicing information that did processor **Mpulses**. Since this difference was observed for processor **Noise32** as well as the higher envelope bandwidth processor **Noise400**, it is difficult to attribute this effect to periodicity cues carried by the amplitude envelope in the range of voice periodicity. To account for the poorer voicing scores from the **Mpulses** processor, it can be postulated that the strong and constant periodic percept that this produces is not readily interpreted by the listener as representing voiceless speech. In contrast, it seems that the strong and constant aperiodic percept from processor **Noise32** can be interpreted as representing voiced speech. That this possible may perhaps be based on our natural experience of whispered speech.

5.4 Results in relation to signals lacking spectral information

In the absence of spectral information, previous studies have shown that pitch information contributes substantially to the audio-visual perception of sentences and CDT (Rosen et al., 1981; Waldstein & Boothroyd, 1994). With just a limited degree of spectral information, only CDT of the speech tests used here shows effects that may be attributable to the presence of pitch information. There is, however, no difference in CDT rates between the fixed-pulse rate processor **Mpulses** and processor **FxNx**, where the carrier for voiced speech carries highly salient pitch cues. The difference that was observed was between noise carrier processors with 32 Hz and 400 Hz

envelope filters, and this effect could also be attributable to the encoding of more rapid spectral changes. Even with CDT, then, there is no clear evidence that pitch information contributes to communication rate in the presence of spectral information, despite the previous findings of strong effects of pitch information when spectral cues are absent.

The identification of consonants from spectrally invariant auditory signals shows also a substantial contribution from the periodicity or aperiodicity of the auditory signal to contrasts of manner and voicing (Faulkner & Rosen, 1999). No such effect is evident here. Where even limited spectral structure is present, it seems that spectral balance differences between voiced and voiceless speech are in themselves sufficient to mark the manner of articulation differences that can also be signalled by temporal cues to periodicity/aperiodicity.

5.5 The role of pitch in speech communication

The present studies are likely to substantially underestimate the contribution of pitch information to communication, especially where paralinguistic cues (e.g. to talker identity) are important. Furthermore, envelope-based pitch cues have been shown to contribute to Chinese sentence perception through similar processors (Fu et al., 1998). The most reasonable interpretation of our findings is not that factors such as voice pitch are unimportant. Rather, we would argue that the intelligibility measures used lack sensitivity to important aspects of speech quality. Since the speech tests used here are essentially the same as those almost universally used in clinical research evaluating cochlear implant benefit, it may be that conventional speech-based measures of benefit are missing aspects of speech perception that are of real importance in speech communication. It is clear too that patients vary greatly in their ability to process pitch cues from CIS stimulation. Furthermore, different processing systems vary considerably in their capacity to transmit pitch-related information to an implant user, in particular where the pulse stimulation rate is too low to allow the accurate sampling of modulations in the voice fundamental frequency range. Since intonation is widely held to be a major factor in the development of spoken language, the role of pitch information in cochlear implant speech processing should not be dismissed simply because it appears to have little impact on intelligibility.

6. Acknowledgements

Supported by a Wellcome Trust Summer Vacation Scholarship to Clare Smith, Wellcome Trust Grant 046823/z/96, and CEC TIDE project OSCAR (TP 1217).

7. References

Breeuwer, M., & Plomp, R. (**1986**). "Speech reading supplemented with auditorily presented speech parameters," Journal of the Acoustical Society of America, **79**, 481-499.

DeFilippo, C. L., & Scott, B. L. (**1978**). "A method for training and evaluation of the reception of on-going speech," Journal of the Acoustical Society of America, **63**, 1186-1192.

Dorman, M. F., Loizou, P. C., & Rainey, D. (**1997a**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,," Journal of the Acoustical Society of America, **102**, 2993-2996.

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

Dudley, H. (**1939**). "The vocoder," Bell Labs. Record, **17**, 122-126.

Faulkner, A., Potter, C., Ball, G., & Rosen, S. (**1989**). "Audiovisual speech perception of intervocalic consonants with auditory voicing and voiced/voiceless speech pattern presentation," Speech, Hearing and Language, Work in progress, University College London, Department of Phonetics and Linguistics, **3**, 85-106.

Faulkner, A., & Rosen, S. (**1999**). "Contributions of temporal encodings of voicing, voicelessness, fundamental frequency and amplitude variation in audio-visual and auditory speech perception," Journal of the Acoustical Society of America, **106**, 2063-2073.

Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., & Rosen, S. (**1993**). "Lip-reading the BKB sentence lists; corrections for list and practice effects," British Journal of Audiology, **27**, 233-246.

Fourcin, A. J., Rosen, S. M., Moore, B. C. J., Douek, E. E., Clarke, G. P., Dodson, H., & Bannister, L. H. (**1979**). "External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings," British Journal of Audiology, **13**, 85-107.

Fu, Q.-J., & Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," Journal of the Acoustical Society of America, **107**, 589-597.

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

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (**1985**). "The contribution of fundamental frequency, amplitude envelope and voicing duration cues to speechreading in normal-hearing subjects," Journal of the Acoustical Society of America, **77**, 671-677.

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.

Group, E. (**1986**). *The BKB (Bamford-Kowal-Bench) standard sentence lists* [Video recordings]. London: Department of Phonetics and Linguistics, University College London.

McDermott, H. J., & McKay, C. M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," Journal of the Acoustical Society of America, **101**, 1622-1631.

Miller, G. A., & Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," Journal of the Acoustical Society of America, **27**, 338-352.

Pollack, I. (**1969**). "Periodicity pitch for white noise - fact or artefact.," Journal of the Acoustical Society of America, **45**, 237-238.

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, Department of Phonetics and Linguistics, University College London, **10**, 1-15.

Rosen, S., Faulkner, A., & Wilkinson, L. (**1999**). "Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants," Journal of the Acoustical Society of America, **106**, 3629-3636.

Rosen, S., Fourcin, A. J., & Moore, B. C. J. (**1981**). "Voice pitch as an aid to lipreading," Nature, **291**, 150-152.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (**1995**). "Speech Recognition with Primarily Temporal Cues," Science, **270**, 303-304.

Shannon, R. V., Zeng, F.-G., & Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," Journal of the Acoustical Society of America, **104**, 2467-2476.

Waldstein, R. S., & Boothroyd, A. (**1994**). "Speechreading enhancement using a sinusoidal substitute for voice fundamental frequency," Speech Communication, **14**, 303-312.

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., & Rabinowitz, W. (**1991**). "Better speech recognition with cochlear implants," Nature, **352**, 2.

Wilson, B., Zerbi, M., Finley, C., Lawson, D., & van den Honert, C. (1997). *Eighth Quarterly Progress Report, May 1 through July 31, 1997. NIH Project N01-DC-5-2103: Speech Processors for Auditory Prostheses.* : Research Triangle Institute.

---

[i] Here and elsewhere, F tests on factors with df > 1 are based on the use of Huynh-Feldt Epsilon correction factors.