IOP Publishing Journal of Neural Engineering
J. Neural Eng. **4** (2007) S58–S71 doi:10.1088/1741-2560/4/1/S08

Simulating auditory and visual sensorineural prostheses: a comparative review

L E Hallum[1], G Dagnelie[2], G J Suaning[1,3] and N H Lovell[1,4]

[1] Graduate School of Biomedical Engineering, University of New South Wales, Sydney 2052, Australia
[2] Wilmer Ophthalmological Institute, Johns Hopkins University, Baltimore, MD, USA
[3] School of Engineering, University of Newcastle, Callaghan, Australia
[4] National ICT Australia, Eveleigh, Australia

E-mail: N.Lovell@unsw.edu.au

Received 10 October 2006
Accepted for publication 22 January 2007
Published 20 February 2007
Online at stacks.iop.org/JNE/4/S58

Abstract
Microelectronic vision prosthesis proposes to render luminous spots (so-called phosphenes) in the visual field of the otherwise blind subject by way of an implanted array of stimulating electrodes, and in doing so restore some spatial vision. There are now many research teams worldwide working towards a therapeutic device, analogous to the cochlear implant, for the profoundly blind. Despite the similarities between the cochlear implant and vision prostheses, there are few instances in the literature where the two approaches are compared and contrasted with a mind to informing the science and engineering of the latter. This is the focus of the present review; specifically, our interest is psychophysics and signal processing. Firstly, we examine the cochlear implant, and review a handful of psychophysical work: the acoustic simulation of cochlear implants and the method used. We focus on the use of normally hearing subjects (played coloured noise bands or sine waves) as a means of investigating cochlear-implant efficacy and speech processing algorithms. These results provide guidance to vision researchers, for they address the interpretation of simulation data, and flag key areas, such as 'artificial' perception in the presence of noise, that require experimental work in coming years. Secondly, we provide an up-to-date review of the body of analogous psychophysical work: the visual simulation, involving normal observers, of microelectronic vision prosthesis. These simulations allow predictions as to the likely clinical efficacy of the prosthesis; indeed, results to date suggest that a number on the order of 100 implanted electrodes will afford subjects mobility and recognition of faces (and other complex stimuli), while even fewer electrodes facilitate reading printed text and very simple visuomanual tasks. Further, the simulations allow investigations of image and signal processing strategies, plus they provide researchers in the field, and other interested persons, a perceptual experience that approximates what a prosthesis will likely afford implantees.

The cochlear implant (CI) and the microelectronic retinal prosthesis, as it is envisioned, are conceptually similar in many ways. Both implant a relatively small number of electrodes at the site of sensorineural elements—hair cells or photoreceptors—and seek to replace normal physiological function via extracellular electrical stimulation of more proximal neural elements. This effectively activates populations of fibres forming the afferent nerves—auditory or optic. Despite the similarities, and despite the clinical success of the CI, there are few instances in the literature where CIs and retinal prosthetics are explicitly compared and contrasted with a mind to informing the science and engineering of the latter. Therefore, the present review is aimed at the vision researcher with an interest in hearing research, whose endeavours in the field of prosthetics may benefit from a better understanding of the cochlear implant. Specifically, our interest lies with

simulating auditory and visual sensorineural prostheses, that is, activation of the sensory epithelium—the cochlea or the retina—in normal subjects by native means—acoustic or visual—in a way that approximates electrical activation. These simulations yield psychophysical data that stand to inform the design of speech or image processing strategies and electrode array designs. The use of normal listeners or observers as opposed to implanted subjects allows the separation of confounding factors, e.g., duration of experience with a particular processing strategy or viability and variability of degenerate neural tissue.

This paper begins with the cochlear implant. Prior to reviewing a handful of studies that, we hope, will provide some direction to future visual modelling studies, we briefly canvas the neurophysiology of the cochlea, speech processing strategies implemented in CIs and the method involved in acoustic modelling. We then examine microelectronic retinal prosthesis and visual modelling of electrical retinal stimulation. We conclude with a discussion of the shortcomings of prosthesis modelling with a mind to informing the interpretation of visual modelling data and the direction of future work in this area.

1. Auditory prosthesis

The CI was first made commercially available in 1982 to those profoundly deaf through hair cell loss. It is estimated that, at present, there exist 85 000 implant recipients worldwide [1]. The improvement in clinical outcomes in recent years (for example, monosyllabic word recognition has improved, roughly, from 10% to 45% between the 1980s and 1999 [2]) is attributed equally to improved speech processing strategies, revised implant candidacy criteria and unidentified factors [2]. This speech processing result suggests that image processing may play a central role in achieving favourable clinical outcomes for a retinal implant. Since this review is aimed at the vision researcher, prior to examining acoustic models of electrical cochlear stimulation we provide here a brief overview of the neurophysiology of the peripheral auditory system (for more detail the reader is directed to [3–5]), speech processing and electrode design.

1.1. Overview of auditory neurophysiology and cochlear-implant speech processing and electrode design

The human cochlea is a fluid-filled cavity in the temporal bone of the skull that analyses sound over approximately ten octaves—from 20 Hz to 20 kHz. The cavity is approximately 35 mm long and is coiled like a snail's shell, spanning approximately 2.5 revolutions, about the auditory nerve which runs through its bony centre—the modiolus. To a first approximation, when the cochlea is viewed in section an upper duct (scala vestibuli) and a lower duct (scala tympani) are apparent, separated by the basilar membrane which is tensioned horizontally across the cochlea. The organ of Corti lies on the basilar membrane and comprises the sensory epithelium—approximately 15 500 outer and inner hair cells, the latter of which predominantly synapse on spiral ganglia forming the auditory nerve (by way of the inner wall of the spiralling cochlea) and convey sound information to the brain. To the vision researcher there are obvious parallels between hair cells and photoreceptors and between spiral and retinal ganglion cells.

The cochlea (contained in the inner ear) receives sound information from the outer and middle ears by way of the oval and round windows (which access the scala vestibuli and scala tympani, respectively), the ossicles (incus, malleus and stapes) and the tympanic membrane. By way of this structure, sound affects a pressure field in the cochlea and generates a travelling wave in the basilar membrane. The pattern of disturbance of the membrane is a function of the incident sound; the membrane is tonotopically mapped (approximately logarithmically) with high frequencies innervating hair cells at the cochlea's basal end (closer to the middle ear) and low frequencies hair cells at the apical end. Although early theories had it that the cochlea functioned like a Fourier analyser (a bank of uncoupled filters each receiving identical inputs [6]), the issue is now understood to be somewhat more complicated: fluid-mediated coupling between portions of the basilar membrane, lateral (neural) inhibition and local feedback mechanisms involving motile outer hair cells all contribute to the analysis.

Of course, there are numerous aetiologies of hearing loss. A subset of the severely to profoundly deaf suffer sensorineural loss which most commonly involves the selective degeneration of hair cells (leaving the auditory nerve and proximal auditory centres intact) and are therefore candidates for a cochlear implant[5]. In the typical multiple-channel implant, an array of stimulating electrodes (between 6 and 22) is implanted in the scala tympani [1, 8]. The array is inserted approximately 22–30 mm deep via a 1–2 mm cochleostomy drilled proximal to the round window. Ideally, so as to minimize the interaction of charge injected via different electrodes and to lower perceptual thresholds, the electrode is positioned perimodiolarly, that is, close to the inner wall of the spiralling cochlea, in close apposition to spiral ganglia. The tonotopic organization of the cochlea is thus exploited along the length of the array. The array is connected to an implanted receiver which, via trans-cutaneous radio frequency transmission, is coupled to an external microphone, speech processor and transmitter worn at the ear.

It is the stimulation strategies implemented by the speech processor that concern the translation of an acoustic signal to injected charge(s) via one or more electrodes. There exist numerous strategies; commercially available devices implement at least one of these strategies according to parameters established in post-surgical, psychophysical testing conducted by the consulting specialist. There exists no universal preference amongst implantees for any one strategy in particular (see, e.g., [9]). Roughly speaking, stimulation strategies may be categorized as sequential or simultaneous, pulsatile or analogue. The former category concerns the number of electrodes that are active at a single point in time; if that number cannot exceed one, the strategy is sequential,

[5] For the relaxation in recent years of implant candidacy criteria concomitant with improved clinical outcomes see [7].

otherwise simultaneous. The latter category concerns the waveforms of injected charge; 'pulsatile' refers typically to bi-phasic, charge-balanced waveforms, the amplitudes of which are modulated between perceptual threshold and comfortable loudness (so-called T and C levels) accordingly. 'Analogue' refers to waveforms that more closely mimic the acoustic waveform presenting at the microphone[6]. Despite there being numerous experimental protocols, those of major clinical significance include SPEAK (spectral peak strategy) [11], CIS (continuous interleaved sampling) [12], ACE (advanced combination encoder) [13] and SAS (simultaneous analogue stimulation; see [14]).

Common to all these strategies is an initial stage of bandpass filters that parcels up some amount of the frequency spectrum between 100 and 10 000 Hz. The output at each filter then undergoes temporal envelope detection (full-wave rectification and low-pass filtering). This is depicted in figure 1. Subsequent to temporal envelope extraction, strategies differ; it is convenient to think of these differences as contingent upon the trade-off between spectral resolution and temporal resolution and the way in which this trade-off is managed. For example, consider implanted hardware capable of delivering 14 400 charge-balanced (sequential) pulses per second (pps). If a stimulation cycle involves only, say, two electrodes, then, by way of Shannon's sampling theorem, the stimulation pattern at each electrode may convey temporal information in the envelope up to approximately 3500 Hz. If a stimulation cycle involves ten electrodes, this rate is reduced by a factor of 5.[7]

The differences between the pulsatile strategies CIS and SPEAK are subtle. In any given stimulation cycle, both strategies sequentially activate between six and eight monopolar electrodes in tonotopic order. CIS, however, provides better temporal resolution (approximately 800 pulses per second per electrode (pps/el) versus approximately 250), though SPEAK provides better spectral resolution (SPEAK involves as many bandpass filters as there exist stimulating electrodes—as many as 22 in some devices; CIS is designed for fewer electrodes—typically six driven by a bank of six filters). Whilst under CIS the stimulation electrodes are 'fixed', under SPEAK the stimulation electrodes are 'dynamic'—for a given cycle, activation occurs at those six or eight electrodes for which filter output is maximum.

CIS was originally proposed as an improved alternative to the compressed analogue strategy [17]. That strategy simultaneously activated multiple monopolar electrodes with analogue waveforms. Such activations are thought to give rise to vector summation in tissue of injected charge and cross-talk between active electrodes, both of which are hypothesized to have deleterious effects on speech understanding (see [18]). SPEAK was originally proposed as an improved alternative to formant-extracting[8] algorithms implemented on small numbers of electrodes [20]; due to very little redundancy being contained in the processed signal, such strategies suffered in the presence of noise (see [20]).

It is convenient to think of the ACE strategy as a CIS–SPEAK union. ACE may provide a simple implementation of either SPEAK or CIS. Further, the number of electrodes activated in any cycle is selectable; the outputs of adjacent filters may be summed when fewer stimulating electrode are employed. The mean stimulation rate (with programmable jitter[9]) varies between 250 and 2400 pps/el, the limit being a maximum of 14 400 pulses within a cycle.

1.2. Acoustical models of auditory sensorineural prostheses

Electrical stimulation of the cochlea can be modelled acoustically. That is to say, experiments may be run involving normally hearing (NH) listeners played sounds, via loudspeakers or headphones, that approximate the perceptual experiences of CI subjects. Acoustic modelling has considerable predictive power regarding outcomes in CI subjects (of which more later), allowing alternative speech processing strategies and electrode configurations to be readily investigated. By way of describing the typical set-up and measures used in acoustic simulations, in this section we discuss in some detail the methods and measures used in a study by Shannon and colleagues [21], which is largely representative of the (vast) literature.

In eight NH listeners, Shannon and colleagues demonstrated near perfect speech understanding (in the absence of competing noise) when an acoustic model of a four-channel implant was used to represent sounds. The simulated sounds were generated as follows:

(1) The original speech signal was digitized at 10 kHz and passed by a high-pass, pre-emphasis filter (cut-off 1200 Hz).
(2) The signal was then presented to a four-filter bank with centre frequencies 0 Hz, 1150 Hz, 2000 Hz and 3250 Hz. Note that the passbands overlapped in the frequency domain at those points where the gain was 15 dB down from unity.
(3) From each filtered signal the temporal envelope was then extracted. This involves full-wave rectification followed by low-pass filtering with cut-off 50 Hz.
(4) These four 'envelope' signals were then used to modulate the amplitudes of four noise signals. Each of these noise signals was previously coloured by one of the passbands from step (2), and each pertained to a particular envelope signal.
(5) Finally, the four signals in (4) were low-pass filtered (0–4000 Hz), summed, the root-mean square was equated to that of the original signal, and the processed signal was played over headphones at a comfortable level.

[6] More detail regarding functional electrical stimulation—e.g. the requirement of charge balance, and the interplay between stimulation rates and refractory periods in target tissue—may be found in [10].

[7] The issue of stimulation rates is somewhat complicated and controversial. Recent work suggests that pulse rates beyond the relative refractory period, even, of auditory neurons may be effective in lowering perceptual thresholds (see [15] and contrarily [16]).

[8] *formant n.* A peak in the spectrum of frequencies of a specific speech sound, analogous to the fundamental frequency or one of the overtones of a musical tone, which helps to give the speech sound its distinctive sound quality or timbre [19].

[9] Fixed-rate pulsatile stimulation may induce perceived 'buzz' in the cochlear-implant subject [11].

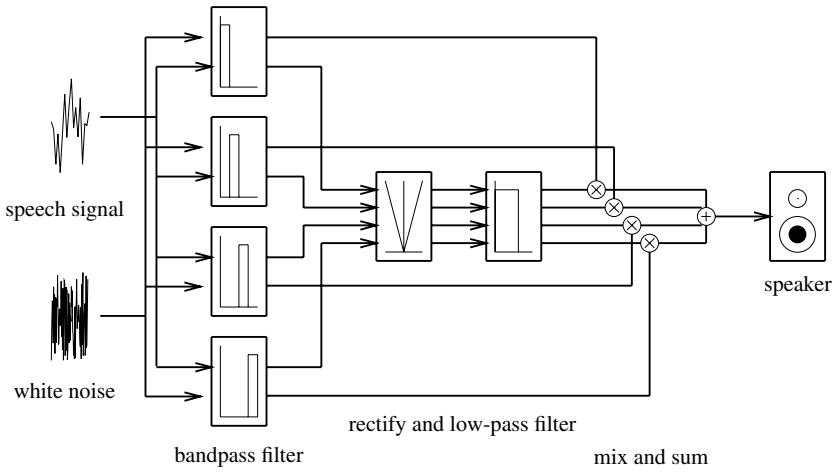

Figure 1. Functional block diagram for the acoustic simulation of a four-channel cochlear implant. In each of four bands (which together parcel up the audible spectrum), the temporal envelope is extracted from the speech signal (full-wave rectification and low-pass filtering) and then mixed with coloured noise. The resulting signal is played to a normally hearing listener via a speaker or headphones.

This process preserves temporal envelope characteristics—those characteristics transmitted to an implantee via the device (see figure 1).

The pre-recorded, spoken test material presented to NH listeners was drawn from 16 consonants presented as a/C/a, that is, spoken between two vowels 'a' (e.g., 'aba'), and eight vowels (h/V/d; e.g. 'had')[10], each listener being required to identify the consonant or vowel from a complete list. Also, a standard battery of spoken, everyday sentences for the assessment of profound hearing loss was presented, with subjects required to repeat back as many words from a sentence as possible. Training prior to testing was allowed until performance stabilized—typically 8–10 h.

From the data, two measures were derived: the articulation scores, that is, the percentage of spoken words that the listener heard correctly, for consonants, vowels and sentence key words; and the information received for consonants. This latter measure is based on the confusion matrix of Miller and Nicely [23]. In a confusion matrix, the spoken consonant is listed along the first column at the left, and the consonant heard by the subject along the first row at the top; cells contain tallies of stimulus–response pairs. For example, if /m/ is spoken but the subject hears /n/, that is, the two nasal consonants are confused, the cell (/m/,/n/) is incremented. As such, the articular score for the entire test, a single statistic, is obtained by summation along the main diagonal. Further, stimulus–response pairs may be grouped by features of articulation: voicing, manner and place[11]. These features may be thought of as the canonical elements for speech discrimination. In effect, they characterize approximately statistically independent communication channels; whilst a listener may typically confuse one feature, the others usually go unaffected. As Miller and Nicely noted, this separation of a complex channel into constituent channels has 'considerable value for the diagnosis both of inefficient equipments and hard-of-hearing people' (p 351) and therefore for rehabilitation techniques.

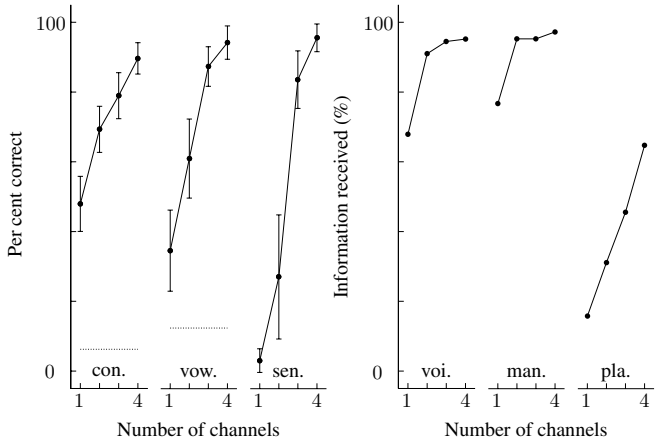

Figure 2. Results for speech perception from the Shannon *et al* [21] study. Generally, speech perception improves monotonically with increasing channel numbers from one to four. Four channels allow near perfect consonant, vowel and sentence recognition, plus the identification of voicing and manner of articulation. The measure 'information received' is based on the normalized covariance of stimulus–response pairs (see [23, 24] for details). Dotted lines denote chance performance. (con.: consonants; vow.: vowels; sen.: sentences; voi.: voicing; man.: manner; pla.: place; data taken from [21].)

Figure 2 shows the results from the Shannon study. Further to the four channels, and 50 Hz low-pass filtering, described in the above five steps, one, two and three channels (wherein the four analysis bands were combined accordingly) and 16, 160 and 500 Hz low-pass filtering of the temporal envelope were tested.

An alternative to the 'noise-band' vocoder (as described above) is the 'sine-wave' vocoder, the difference being that

[10] The reader unfamiliar with phonetic notation is referred to [22, chapter 2].

[11] A voiced consonant involves the vibration of the vocal cords, e.g., /b/ versus the unvoiced /p/. Manner involves the way in which the tongue, lips, and other speech organs involved in articulation make contact in the production of sound. Place of articulation refers the whereabouts in the mouth constriction (for the most part) occurs, e.g., /p/, /t/, and /k/ may be classified as 'front', 'middle', and 'back' respectively.

sinusoidal tones as opposed to bands of noise are played to NH listeners. Dorman and colleagues [25] compared the two approaches and demonstrated little difference in vowel, consonant and sentence results. By contrast, in a recent paper Gonzalez and Oliver [26] showed that NH listeners played sounds via the sine-wave vocoder better identified gender and speaker, especially where only a few channels (between three and five) are concerned. It is widely held that both methods are good for prediction of CI subject performance (of which more later), however, as to why there exist subtle, task-dependent differences between the noise-band and the sine-wave vocoders remains an open question [26].

1.3. Some results concerning acoustic modelling of electrical stimulation of the cochlea

There are presently no data to suggest that visual models of electrical stimulation of the retina (discussed subsequently) are good predictors of clinical outcomes. Verification to this end should become a priority for visual modelling workers once more clinical data are available; at the time of writing, several clinical trials of retinal stimulators are underway [27–29]. However, vision researchers can draw encouragement from the success of acoustic modelling of electrical stimulation of the cochlea in predicting performance in CI wearers, although the data require some interpretation.

Dorman and colleagues [30, 31] have shown that some CI listeners are able to extract as much speech information as NH listeners afforded a simulation with equivalent channel numbers. Their cohort of ten NH listeners averaged at least 80% correct identification of vowels; the performance of four of seven CI listeners fell within one standard deviation (s.d.) of this result. The NH cohort averaged 85% transfer of information [23] for place of articulation of consonants; five of seven implantees fell within one s.d. Of note is the fact that NH listeners were unpracticed, whereas CI listeners had between one month and four years' experience with the speech processing strategy tested. Similar results were observed by Dorman and colleagues in the presence of speech-shaped noise. Further, they observed an apparent increased susceptibility of the CI listener to noise as compared with the NH listener. Vowel recognition decreased (from approximately 80% to 70%) in the NH cohort as the signal-to-noise ratio (SNR) decreased (from 15 to 5 dB); these decreases, however, were out-stripped by CI listener performance, where four of seven CI listeners fell within one s.d. for the 15 and 10 dB conditions, but only three of seven for the 5 dB condition.

Friesen *et al* [32] also tested susceptibility to noise in five NH listeners and 19 CI listeners—figure 3 is representative of their data. In quiet, and for modest amounts of noise, note from figure 3 how, for lower channel numbers (<10), the mean NH listener performance predicts that of the best performing CI listeners. The Dorman *et al* and Friesen *et al* studies are typical of acoustical modelling studies in that they demonstrate how the performance of NH listeners apparently forms an 'upper bound' for the prediction of CI listener outcomes. Friesen *et al*, however, note that an increasing amount of noise tends to even up the NH versus CI comparison, although NH listener

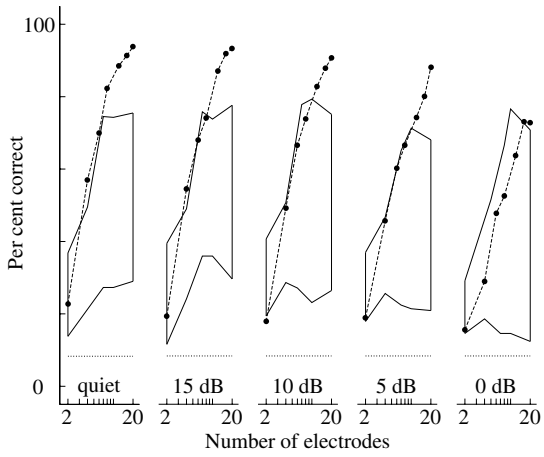

Figure 3. The Friesen *et al* [32] data for vowel recognition (presented as h/V/d) for increasing amounts of competing noise (signal-to-noise ratio decreases from left to right as noted). For each noise condition, the area enclosed by the solid lines shows the range of performance of the CI cohort; the filled circles depict the mean NH cohort performance. The dotted lines denote chance performance (8.3%). (Figure after [32].)

performance is better than CI listener performance on all tasks; in the '0 dB' condition (rightmost panel), the best CI listener outperforms the mean NH listener. This was also the case for monosyllabic word recognition. Friesen and colleagues proposed that, for noisy conditions, the extra practice afforded CI listeners (in their day-to-day use of the device) in extracting information from severely degraded signals may explain their relatively good performance.

It is interesting to note that CI listener performance improves only to about eight or ten electrodes, a finding replicated in other cohorts (for example, [33]). By contrast, NH listener performance generally shows little sign of asymptoting over a range of 2–20 channels. This indicates that, for many electrodes, say 20, the information transmitted by the electrode array to the CI listener is only partially received, and that increased effective channel numbers in CI listeners, as opposed to increased electrode numbers, would have marked effects for speech understanding. The vision researcher working on prosthesis development would do well to be mindful of this transmission–reception gap, especially where implantable array manufacture is concerned; unless information received, borne out by psychophysical studies, increases with the number of implanted electrodes, high-density arrays [34] are likely to confer little benefit. Numerous factors are hypothesized to contribute to CI listeners' performances asymptoting at a relatively small number of channels [30], including (1) the nonuniform survival of spiral ganglia in CI listeners, whereas NH listeners presumably have access to a wholly healthy inner ear; (2) the interaction of current in tissue; (3) shallow electrode insertion; (4) poor resolution of intensity differences and (5) poor resolution of the temporal waveform. Indeed, the use of acoustic models, and the contrast of those outcomes with CI listener data, is an ongoing means of characterizing this transmission–reception gap. One would anticipate an

analogous gap in retinal implantees, since the five factors mentioned also apply in retinal tissue.

In light of the above acoustic studies, it is interesting to note that the robustness of prosthetic visual perception in the face of noise has yet to be determined. Despite the fact that signal-to-noise ratio could be readily and systematically varied in a cohort of normal observers, it is as yet unknown how noise will affect perception. Not only might the addition of noise make for improved visual modelling (of which more later) if, like the Friesen *et al* data, increased noise makes for more comparable performances between the CI and NH cohorts, but these data would inform image processing (the pre-sample filtering), since image processing approaches vary in noise tolerance from one to the next. The data may also inform post-implantation psychophysical device fitting. The fitting process potentially introduces a major source of noise: sampling jitter. That is to say, in the fitting of a device, the subject is required to indicate those locations in the visual field occupied by phosphenes (rendering locations, probably by way of a Humphrey field analyser); these data are central to the image processing (since they determine sampling locations in the underlying image) and are subject to error in their acquisition.

If, for example, the sampling jitter, considered over the entire phosphene image, describes a bivariate normal, the resulting phosphene image would incur a noise that increased monotonically with spatial frequency in the underlying image, that is, fine details would be effectively viewed in the presence of noise [35]. The question then arises, to what extent can the observer of the phosphene image develop internal models that compensate for this noise? With this in mind, there is an analogous body of acoustic modelling work that is of interest—work concerning speech recognition where either (1) noise (typically speech-shaped and temporally modulated) is mixed with the acoustic signal presented to NH listeners or (2) spectral distortion is introduced, that is, a situation where incoming acoustic signals ultimately activate cochlear places other than those affected under normal, biological operation of the cochlea. As noted by Dorman *et al* [36] (discussed below), CI patients may be able 'to compensate for only distortions of a modest magnitude' (p 2996), and this learned compensation may account for CI-processed speech sounding abnormal after initial activation of the device (according to subjective reports), but after some use its becoming more intelligible. The analogy is discussed further in the following paragraphs.

The CI involves spectral distortion, ipso facto. That is, the incoming acoustic signal ultimately activates cochlear places other than those affected under normal, biological operation. This is due to the fact that CIs are designed with Greenwood's [37] frequency-cochlear place equation in mind, and speech processor filters offer only limited spectral programmability, though there exist many between-patient variables, e.g., electrode insertion depth and interaction of charge in viable peripheral neural tissue. Therefore, understanding the responses of central auditory templates to spectral distortion, and their robustness in the face thereof, is important for the improvement of speech processing strategies. As discussed in [38], many psychophysical data suggest that auditory pattern recognition is neither 'positionally relocatable' nor does it simply encode relative positions of spectral features. Therefore monotonic spectral shifts, plus spectral expansion and compression (characteristic of the CI), yield adverse effects for speech recognition. The correct identification of vowels, for example, is heavily dependent on spectral cues that typically exist at low and middle frequencies [39] and therefore typically activate specific (tonotopic) neural substrates. In the Nucleus-22 implant implementing SPEAK, for example, an acoustic range of 150–10 kHz, which usually occupies a 25 mm segment of the cochlea, is mapped to a 16.5 mm segment. This spectral compression aside, the position of the segment in question varies with insertion depth.

Dorman and colleagues [36] used an acoustic model to simulate shallow insertions (22–24 mm cf a 'normal' 26 mm insertion) of a five-channel, CIS device with inter-electrode spacing of 4 mm. This was achieved by mismatching analysis frequencies and carrier frequencies of sine waves comprising the vocoder; e.g., in the 22 mm depth condition, the sine wave simulating the apical-most electrode was of frequency 831 Hz, though the analysis filter driving the intensity of this carrier passed a narrow band of the original signal centred at 418 Hz. Subjects (nine NH listeners with 12–15 hours' practice) demonstrated a main effect for insertion depth on speech understanding; generally, the 'normal' and 25 mm conditions produced best results and were significantly different from 22–24 mm insertions. It is interesting to note that for both the normal and 25 mm depth conditions, the sine waves in question lay within the pertinent analysis bands; Dorman and colleagues hypothesized that in similar cases of only slight analysis-carrier mismatch speech recognition may go unaffected. This nonlinearity may likewise be manifest with regards electrical stimulation of the retina. It is interesting to note the analogy that exists with visual modelling studies involving eccentric viewing of stimuli (discussed subsequently)—the changed preferred retinal locus requires some learning by observers. As Dorman and colleagues noted, the effect of electrode insertion depth is difficult to discern in CI subjects, where many factors are confounded; a patient with deep insertion may suffer a paucity of viable neural tissue near the active electrodes, while a patient with a short insertion may have an abundance of stimulatable neural elements near each electrode. Hence, NH listeners and an acoustic model are well suited to the specific aims of this study. Other factors that may vary across a cohort of CI listeners include, for example, demographics or duration of experience with one or other clinical speech processing strategy. Similar advantages exist with regard to normal observers and visual modelling as compared with experimenting with implanted subjects.

Another focus of visual and acoustic modelling researchers alike is the development of simulations that better approximate the perceptual experiences in normals to those of subjects of electrical stimulation. In this regard, Fu and Nogaki [40] adapted, apparently with some success, the noise-band vocoder in an attempt to model between-electrode interaction of charge in the basilar membrane. In their adaptation, rather than noise bands falling off in the frequency domain at 24 dB/octave, the bands were smeared

(6 dB/octave fall-off) so as to overlap. Smearing (or otherwise), channel numbers (4, 8, 16) and the temporal characteristics of a temporally modulated masker (a model of competing speech) were then parametrically varied in six NH listeners, and their speech understanding was compared to ten CI listeners—'good users' of the device. The mean performance of CI listeners approximated that of NH listeners afforded four smeared (6 dB/octave) channels. The best CI listeners' performances approximated those of NH listeners afforded 8 or 16 smeared channels or four unsmeared channels (24 dB/octave). Further, the smearing model apparently predicted CI listeners' release from masking, that is, their capacity for speech understanding despite the noise. CI listeners' speech understanding deteriorates quickly in the face of noise, especially temporally modulated noise and competing speech. NH listeners afforded smeared channels only achieved release from masking for 16 channels (whereas no-smearing NH listeners apparently achieved some release from masking for both 8 and 16 channels). In general, CI listeners demonstrate a marked susceptibility to noise; as the intensity of competing speech increases, speech understanding rapidly diminishes. Therefore, understanding CI listeners' patterns of comprehension when faced with noise remains a major challenge presenting to those interested in designing more effective speech processing strategies and electrode designs. Thus far, very few visual prosthesis simulations have paid any attention to the role of noise; all the work cited above suggests that adding noise to the simulations will provide a more realistic impression of what the degenerated retina will convey in response to electrical stimulation. Hence, the addition of noise may be central to the development of better visual models.

Another parameter that goes largely untested in the visual modelling literature is the effect of quantization (the number of grey levels or sizes to which phosphenes can be modulated). By way of analogy, Loizou and colleagues [41] examined the effect of amplitude quantization, wherein the temporal envelope (as shown in figure 1) was quantized between threshold and comfort to some number of levels—2 through 512—thus affecting, for the most part, the temporal information conveyed to the listener. Quantized signals were presented to both CI listeners for the identification of consonants (the recognition of which is well known to be largely dependent on temporal cues) and NH listeners for the identification of sentences and monosyllabic words. The CI listeners (five- and six-channel, CIS devices), all of whom had at least three years' experience with the present speech processing strategy, achieved asymptote performance (ranging between 45% and 90% correct identification) with as few as eight levels of quantization. This indicates that the performance of CI listeners is primarily limited by the number of channels, as opposed to the number of quantization levels, since, clinically, most patients are afforded at least eight discriminable levels. NH listeners, afforded a short practice period and presented signals from a sine-wave vocoder, were used in an attempt to separate interaction between number of channels and number of quantization levels. Generally, as the number of channels increased, the number of requisite levels decreased from either 8 or 16 to 4. As expected, vowel recognition was more robust, as compared with consonants, when temporal information was degraded; for all but the six-channel condition, asymptote was reached at four levels, whereas most consonant recognition required eight levels. For consonants processed via a six-channel sine-wave vocoder, NH listeners' performances asymptoted at eight levels, consistent with results in CI listeners. Further, feature analysis [23] suggests that eight levels are required for the reliable recognition of place, manner and voicing.

Taken together with the results of Drullman [42], wherein NH listeners demonstrated high speech intelligibility with two quantization levels and 24 channels, the result indicates trade-off between spectral resolution and amplitude resolution of the temporal envelope; fine temporal envelope cues are not necessary for high recognition where an increased number of spectral cues exist.

2. Retinal prosthesis

There is a growing body of (multidisciplinary) literature concerning the bioengineering of a retinal prosthesis (for a recent review see Weiland *et al* [43]; see also [44]). The small number of envisioned devices typically feature an implanted array of electrodes in close apposition to the retina, and a unit worn externally to the body. The external unit typically comprises a digital, or infrared, camera and associated signal processing capability which communicates by radio frequency signals with an application-specific integrated circuit connected to the array [45]. For a recent clinical trial of an epi-retinal, 16-electrode array driven per-sclerally by a modified Clarion cochlear implant, see [27]. For the sub-retinal approach, see [46].

The cornerstone of retinal prosthesis is the phosphene—the perceived luminous spot evoked by an electrode stimulating viable layers of the degenerate retina [47–50]. The expectation is that, in the same way that the CI modulates the loudness of a pitch by modulating a stimulus profile at an electrode (of which more later), a clinical retinal prosthesis will modulate the size or intensity of a (localized) phosphene.

2.1. Visual modelling of retinal prosthesis

Electrical stimulation of the retina can be modelled visually [51–64]. That is to say, experiments can be run wherein face recognition, for example, is assessed in normal observers of phosphene images that depict (phosphenized) faces. In such experiments, the phosphene images generate spatial patterns of retinal activation in normal observers that seek to approximate patterns of excitation arising from electrical stimulation of retinal tissue (see, e.g., [65]). Figure 4 depicts a number of phosphene images.

How are phosphene images generated? The underlying image (the top-left panel in figure 4) is sampled at the locations of the phosphenes. Each of these sampled image intensities then modulates the intensity and/or size of a phosphene which occupies some location in the subject's visual field. For example, if the underlying image intensity at coordinates

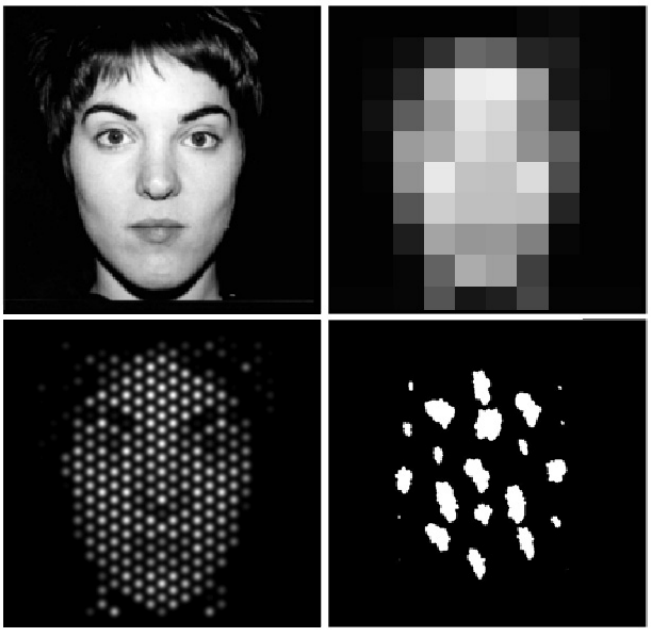

Figure 4. Three phosphene images of the face in the top, leftmost panel (image source: [66]). These show three different means of simulating prosthetic vision: (clockwise from top-right) square, contiguous phosphenes of varying intensity; disordered, high-contrast phosphenes based on acute human trials (image source: [67]); Gaussian spots as phosphenes.

(x_1, y_1) is 75% grey, the phosphene at the corresponding location in the visual field may be modulated to 75% of its maximum size. Prior to sampling, however, pre-filtering of the underlying image occurs—image processing that anti-aliases, to some extent, the resultant phosphene image, and may serve to extract salient features of the underlying image (e.g., see [58, 62]; see also [68] for the application of 'intelligent' algorithms to salient feature extraction). Image samples are then quantized to some number of grey scales or sizes, since implantable neurostimulator designs presently deliver some finite number of pre-programmed, bi-phasic pulses (e.g., see the neurostimulators by Suaning and Lovell [69], based on a 5-bit digital-to-analogue converter or Sivaprakasam *et al* [70] based on a 6-bit converter).

2.2. *Review of visual models of retinal prosthesis*

Much of the visual modelling data concerns reading [51, 54–57, 59]. This is not surprising considering that, according to subjective reports by low vision sufferers [71], the restoration of reading and mobility are likely to confer the most benefit to low vision sufferers in their daily lives. Cha and colleagues undertook a series of three studies [51–53] seeking to quantify, through simulations, the usefulness of an electrode array implanted intracortically in a single gyrus of the primary visual cortex. Although at the time there existed literature concerning perception and 'coarse-quantized' or 'blocked' images (e.g., see [72]), that is, images of much reduced spatial resolution, there arose an anxiety as to whether an array of so few phosphenes (on the order of hundreds) as compared with the number of neural elements comprising the human visual system would have any clinical efficacy (it is interesting to note that a similar anxiety existed with regard to cochlear implantation [73] as cited in [74]). The electrode array of interest to Cha and colleagues measured 1 cm square and comprised some 625 electrodes; given visual cortical magnification in humans (approximately 6 mm/deg), the stimuli used by Cha *et al* involved phosphene images with resolutions 25 × 25 occupying the central 1.7° of the visual field and similar. The more recent simulations—Sommerhalder and colleagues, Fu *et al* and the present authors' groups—have adapted Cha's scales to better suit retinal, as opposed to cortical, prosthesis. Here, the number of phosphenes involved is smaller and the inter-phosphene spacing larger, in accordance with a number of 'retinal' results, the first and foremost being that larger electrodes, as opposed to smaller, are required in order to elicit perception if charge densities at the electrode–tissue interface are to be kept below deleterious levels [10, 49] (see also [65]).

Table 1 summarizes much of the reading data, which, taken together, demonstrate that a low resolution phosphene array may still afford subjects good, albeit slow, text comprehension. Not surprisingly, the most important factor involves the number of phosphenes and the effective sampling density (by comparison, reduced contrast, e.g., has little adverse effect for reading). That is to say, 'zoom' matters; the number of characters displayed on the phosphene array at any one time is a factor with implications for print sizes that will afford implant recipients improved visual function. For example, one can readily envisage an implant that employs digital zoom in the external image processor for reading printed text such that the desired number of characters are presented on the phosphene array at once. The representation of more characters is concomitant with decreased sampling density (and therefore decreased acuity), and the presentation of fewer characters encroaches upon the average reader's ability to assimilate numerous characters at a single glance. For this reason, Hallum *et al* [75] proposed the hexagonal arrangement of phosphenes in the visual field (and accordingly the hexagonal manufacture of implantable arrays) which makes for 14% higher density than the square mosaics used to date by Cha *et al*, among others.

A caveat in interpreting the results from the first two studies by Sommerhalder and colleagues [55, 56] (see table 1) is the 'fixed' nature of phosphenes; their experimental set-up was equivalent to observing a paper-based printout of phosphene images through a restricted window stabilized on the retina. In the third study by Sommerhalder and colleagues [57], phosphenes were stabilized on subjects' retinas, as opposed to simply the viewing window. Thus subjects were allowed to affect interlaced sampling with the phosphene array and temporally integrate phosphene outputs over short time spans, which one would expect to have beneficial effects on performance. Accordingly, the data showed that the requisite resolution of the phosphene array for at least 'good' text comprehension decreased from about 500 phosphenes to about 320. Further, this third study sought more realistic simulation of current spread in the retina; phosphenes took on the appearance of Gaussian spots (see figure 4), as opposed

Table 1. Summary of reading and acuity performance of subjects afforded simulated prosthetic vision. Note that normal observers read at approximately 250 words/min [77] and perfect vision corresponds to 0.0 logMAR.

Study	PI resolution[a]	Performance	ppc and zoom[b]	ES[c] (μm)	Notes
Cha *et al* [51, 52]	32 × 32 (1.7)	0.11[d] } 160–200[e]	} 6 × 6 and 4	17	} Comparable to control reading speeds
	25 × 25 (1.7)	0.18[d] }		22	
	16 × 16 (1.7)	0.48[d]/100[e]		34	
	10 × 10 (1.7)	0.70[d]/50[e]		58	
Fu *et al* [76]	6 × 6 (5.7)	10[e]		290	Error rates as low as 10% despite slow reading speed
Hayes *et al* [59]	16 × 16 (11.8)	1.32[d]/25[e]	} 5 × 5	230	
	6 × 10 (11.3)	1.82[d]/1[e]		600	
	4 × 4 (7.3)	2.0[d]		600	
Sommerhalder *et al* [55]	50 × 17.5 (20.0)	} NPC[f]	} zoom = 4	120	Less visual field occupancy exaggerated
	28.6 × 10 (20.0)			220	adverse effect of eccentric viewing
	50 × 17.5 (20.0)			120	Eccentric viewing (10°)
	28.6 × 10 (20.0)	50%		220	Eccentric viewing (10°)
	20 × 7 (20.0)	<50%		320	Eccentric and central viewing
Sommerhalder *et al* [56]	28.6 × 20 (10.0)	13%–NPC		100	Viewing (15°); performance range: 2 months' practice
Chen *et al* [60]	10 × 10 (13.3)[g]	1.69–1.59[d]		500	Performance range: 10 sessions' practice

[a] Phosphene image resolution. Resolutions are given as *width phosphenes* × *height phosphenes* (*visual field occupancy (width)*). For example, 32 × 24 (1.7) denotes a square mosaic of 768 phosphenes in the central 1.7 × 1.3° of the visual field.
[b] Phosphenes per character (measure of sampling density) and zoom (the number of characters represented in the phosphene image at once).
[c] Electrode spacing. Simulation corresponds to spacing of electrodes implanted at the retina.
[d] logMAR.
[e] Words per minute.
[f] Near perfect comprehension.
[g] Hexagonal phosphene mosaic was used; resolution of equivalent square mosaic is provided for comparison.

to squares of uniform intensity that tiled the retinal locus in question (as per studies I and II). The study found an apparent optimal Gaussian spot size—spots much broader than the effective point sources used by Cha *et al*—though no statistical difference between the Gaussian and square spot conditions was found. This result has implications for the perceptual effects of current spread about stimulating electrodes (see [78, 65]).

In a small cohort ($n = 4$), Thompson *et al* [58] examined face recognition, and further to the above-mentioned parameters, tested the effects of intensity quantization, contrast and phosphene dropout rate (wherein a number of randomly chosen phosphenes was left inactivated). Their results demonstrated 92% accuracy for a 25 × 25 array (subtending the central 15°) and 60% accuracy for a 10 × 10 array (central 6°; results corroborated by Hallum *et al* [67] in a large cohort ($n = 38$)), both significantly above chance. The effects of most factors were as expected: increased array resolution, decreased quantization and decreased phosphene dropout all made for improved face recognition. The effects of the phosphene size (pillbox-shaped) and the intervening gaps, however, were somewhat more complicated: 'narrower gaps seem to yield a slightly better identification than wider gaps . . .. Thus minimizing gaps between electrodes while maintaining separable phosphenes may result in improved performance' (p 5040). Extremes of quantization (e.g., when only two levels of grey were used in rendering the phosphene image), and high phosphene dropout rates (e.g., 70%), challenged subjects, an effect that was exacerbated when the contrast of stimuli was further reduced.

The same laboratory recently published a study on reading performance under conditions very similar to those of the face identification study [63]. Reading speeds of 30–60 words per minute without errors were recorded for some parameter combinations. In general, reading accuracy and speed were influenced by all parameters. Reading accuracy exceeded 90% if the following conditions were met: at least 3 ppc were presented, and dropout did not exceed 50%. Reading speed only deteriorated below 20 words per minute when accuracy fell below 90%; this also happened if the grid spanned less than two characters, especially at low contrast. Grey scale resolution had little effect on reading performance; in fact, at low contrast the lowest grey scale resolution (i.e., two levels) resulted in the best performance, since this effectively maximized the distinction between text and background. Another important finding confirmed in this study was the importance of practice: both reading speed and accuracy increased as the subject read a total of 384 text fragments of 9–12 words each, equally distributed across contrast level and other stimulus conditions; improvements typically occurred across a span of 100 trials for any particular condition, earliest for easy conditions at high contrast, and only towards the end for the hardest conditions at low contrast.

A subset of the Hayes *et al* [59] data concerns symbol identification and the performance of day-to-day tasks. The data suggest that a 6 × 10 array is sufficient for simple

hand–eye coordination tasks, though more complex visuomanual tasks require the resolution of their 16 × 16 array or better. In our laboratory, we have observed that the presence of a subject's own hand in the (phosphenized) field of view, if only because it affords a better sense of scale, makes for a marked improvement in visual cognition. This being the case, it is surprising that there exist so few data regarding visuomanual tasks; this is an area that needs further attention. Moreover, Hayes *et al* reported that square- and house-shaped symbols are more easily recognized on a 4 × 4 mosaic as compared with a circle. This was taken to mean that 'edges and corners are most easily recognized at very low resolution' (p 1023). A confounding factor here, however, is the arrangement of phosphenes. Hayes *et al* used a square mosaic; a hexagonal mosaic, by contrast, is more invariant when rotated and may afford better recognition of, e.g., circular symbols. Chen *et al* (esp. see their figure 13), in comparing hexagonal and rectangular phosphene arrays, showed that acuity is a function of visual meridian, which furthermore varies with phosphene layout. Overall their hexagonal array outperformed the square array (1.54 versus 1.74 logMAR); indeed, a 0.06 logMAR difference would be expected on the basis of sampling density alone. Hayes *et al* [59] measured acuities of 1.96, 1.82 and 1.32 logMAR for the 4 × 4, 6 × 10 and 16 × 16, respectively. Like Chen *et al*, those workers note the importance of scanning for improved acuity. In the Chen *et al* data, the hexagonal array outperformed the square array, a contrast that narrowed with practice, suggesting that subjects adopted effective scanning techniques to overcome the limitations of the square array. Cha *et al* [53] generally found better acuities (ranging from, approximately, 0.70 to 0.10 logMAR), though the arrays they used subtended lower visual angles.

Cha *et al* [52] conducted a mobility study wherein subjects wearing a head-mounted system were required to navigate a maze comprising high-contrast obstacles and other positional cues (white on black background). Here, as field of view (the acceptance angle of the camera), and number of phosphenes comprising the array, increased up to 30° and 25 × 25 (subtending the central 1.7°) respectively, walking speeds (and lack of collisions) became comparable to the control condition (0.83 m s^{-1}). Increases in the number of phosphenes beyond 25 × 25 were of little benefit. Note that the number of phosphenes, as opposed to the inter-phosphene spacing, was much more highly correlated with performance, indicating that acuity, unlike other above-mentioned tasks, is less important where mobility is concerned. The subjects used in this Cha *et al* study were first trained over the course of three weeks, wherein they learned to cope with the perceptual consequences of object minification (wherein a large field of view is represented on the small subtense of the phosphene array), and accordingly increased their walking speed by five fold. Too small a field of view slowed subjects' walking speeds, as they were required to increase the amount of scanning. Inefficient head movements were reported to cause loss of balance.

A recent study at Johns Hopkins [64] examined the performance of four normally sighted and one low vision subjects in an eye–hand coordination task with 6 × 10 Gaussian-shaped phosphenes, in both free-viewing and stabilized conditions. These conditions were chosen to approximate the anticipated geometry of near-term retinal implants such as the Second Sight A60 implant scheduled for clinical trials within the next year. The subjects were required to first count the 1–16 white fields randomly placed in an 8 × 8 checkerboard and later cover each field with a black checker, while inspecting the board with a downward-looking camera mounted on the front of their video headset. Subjects never saw the board in normal view, so all information was obtained from the phosphenized image. As in other trials, practice proved extremely important in this test, albeit mostly in terms of timing: subjects made very few counting and placement errors, but times improved by a factor of 3–5 from early to later sessions.

An important aspect of the Dagnelie study [64] was the role of retinal stabilization of the phosphenized image with the help of an infrared video-based pupil tracker. Without exception subjects responded to stabilization by minimizing their eye movements during trials, rather than trying to use any advantage they might have gained from the inevitable 1–2 frame delay imposed by the eye tracker; in doing so, subjects became very efficient in using head movement to maximize the information gained and with practice achieved performance levels indistinguishable from those in free-viewing trials. This finding is of particular importance in view of the discussion among visual prosthesis developers, whether eye-movement compensation will be required for those designs using a head-mounted camera. The results reported in this study suggest that eye-movement compensation will not be necessary, at least for eye–hand coordination tasks.

The approach of the University of New South Wales–University of Newcastle Laboratory is somewhat different from the above-mentioned studies. While we have sought to quantify the usefulness of prosthetic vision in an acuity task [60], and in a task wherein a small moving target was tracked [62], our main focus is image processing, that is, the pre-processing that lies functionally intermediate to the digital camera that acquires the scene and the phosphene image that is rendered in the visual field. We have demonstrated that non-trivial image processing can improve a subject's performance in both tasks. The Chen *et al* [60] data show that subject acuity is strongly influenced by the pre-processing step; for example, for the square phosphene mosaic used there, if the circular, uniform-intensity kernel that pre-filters the image has width 30% of the inter-phosphene spacing, acuity is improved by approximately 0.12 logMAR as compared with a 90% width. Hallum *et al* [62] showed that if the pre-processing involves a Gaussian blur, as opposed to a uniform-intensity blur (as used in other studies), fixation and pursuit of a small target is improved in accuracy by 35.8% and 6.8%, respectively. These results suggest that, to some extent at least, results from information theory may be applied in an attempt to derive optimal pre-processing, and ultimately improve the vision afforded implantees [79].

3. Sensorineural prosthesis modelling: improvements and interpretations

In section 1.3 we touched on a key issue facing visual modellers: the reconciliation of data from simulations and studies involving retinal implantees. There are presently no data to suggest that visual models of electrical stimulation of the retina are good predictors of clinical outcomes; verification to this end should become a priority for visual modellers once more human-trials data are available. With regard to the cochlear implant, the improvement of acoustic models is ongoing (see, e.g., discussion surrounding Fu and Nogaki in section 1.3). We have begun analogous work to this end [67], based heavily on the human-trials data of Rizzo *et al* [49], that attempts to account for current spread at the vitreoretinal interface, and perceptual effects thereof, by way of irregularly shaped phosphenes.

Another issue is the perception of noisy phosphene images, not only because a device is necessarily subject to noise (e.g., in the implantable microelectronics), but also as a means of developing better visual models (inasmuch as the performance of normal observers, which likely forms an upper bound on implantee performance, would be hindered by the addition of noise and would therefore redress the normal–implantee discrepancy, analogously to the above-mentioned study by Friesen and colleagues). The issue, however, may not be straightforward. Morrone and co-workers [80] showed that the addition of noise to coarse-quantized images paradoxically improves face recognition as it negates the Harmon–Julesz illusion [81]. Also, Hallum *et al* [35, 82] have argued that the addition of noise to phosphene images facilitates the veridical perception of texture.

The stabilization of phosphene images on the retina, which makes for a more accurate physical simulation of retinal prosthesis, wherein the stimulating array is affixed to the retina, is an improvement that has also received only limited attention. To date, most studies have involved computer displays that are freely viewed, that is, the subject's gaze moves relative to the phosphene image. For image stabilization, the phosphene array is scanned over the 'underlying' scene by way of eye movements; for free viewing, some other modality, such as head movements or use of a joystick or computer mouse, is used. Achieving the former in simulations, however, is technically difficult; rigorous monitoring of eye movements requires sampling at rates of at least 200 Hz [83]; the re-rendering of phosphene images at high rates, plus the re-processing of the underlying scene by (non-trivial) image analysis, places extreme demands on the computers involved (see, e.g., Fornos *et al* [57] wherein image rendering was necessarily simplified). The consequences of stabilized versus non-stabilized phosphene images for interpretation of data are as yet largely undetermined. Whilst Cha *et al* [51] examined the differential effects of each in an acuity task and found no effect, further work in this regard is warranted.

In considering the acoustic modelling studies discussed in section 1.3, the question arises, how is the CI listener–NH listener discrepancy explained and how can this inform visual modellers? One explanation involves the pattern of disturbance of the basilar membrane. Acoustic stimulation makes for somewhat complicated disturbance patterns and therefore complicated patterns of innervation of hair cells; for example, see the discussion in [5] surrounding a 3 kHz pure tone that sets up a travelling wave in the membrane, extending approximately one-third of the membrane length, especially in the basal-ward direction from the point of maximum disturbance. By contrast, electrical stimulation of the membrane, in its ideal form at least, is simple; a restricted locus of the membrane is activated as per Greenwood's place-frequency function [37]. That is to say, the fixed-filter set-up implemented by most clinical speech processors (as described above) is more an implementation of the Helmholtzian cochlea than it is the contemporary concept of the cochlea, which includes fluid mechanics and motile outer hair cells. The discrepancy between fibre activation arising from acoustic and electric stimulation of the cochlea is further exacerbated by the largely unknown dynamics of current spread through tissue in the inner ear. This highlights the relevance of biophysical and systems neuroscience models of electrical retinal stimulation to visual modelling, such as those of Cottaris and Elfar [65] and Dokos *et al* [78], which may be used to quantitatively inform some of the more intractable perceptual phenomena that accompany phosphenes, e.g., shape and textural irregularity, non-repeatability, temporal flicker and movement sensation. By the nature of their being perceptual, these phenomena highlight a major difficultly in modelling sensorineural prosthetics.

Another explanation for the CI listener–NH listener discrepancy involves NH listeners' perception being informed by a healthy inner ear. The CI listener, on the other hand, suffers from the degeneration of spiral ganglion cell populations, apparently brought on by the absence of neurotrophic factors normally expressed by (now degenerate) hair cells ([84] and references therein). Further, hair cells are thought to be the source of spontaneous activity in auditory fibres, of which the deafened ear shows a marked paucity, and it is the hair cell synapse that contributes to timing jitter of spiking within individual fibres. Therefore, acoustic stimulation produces stochastic differences in spiking activity between fibres; electrical stimulation makes for synchronization in spiking activity [85, 86]. If electrical stimulation were able to produce spiking activity more like the physiological norm, presumably CI listeners would be afforded greater dynamic range and more orderly loudness growth [86], both of which are readily exploited in NH listeners. In this connection, there is some interest in high-frequency stimulation of the cochlear so as to induce between-fibre spiking asynchrony. Future work could determine whether high-frequency stimulation is relevant to retinal prosthesis, since, analogously, outer-retinal degenerates suffer atrophy of the retinal ganglion cell later.

Finally, it is of interest to visual modellers that, despite more than two decades' development of acoustic models of cochlear stimulation, results still come with caveats. In this connection, the reader is directed to [18], where it is asserted that acoustic modelling results 'should be interpreted in terms of the trends that are predicted rather than as a quantitative

estimate of cochlear-implant subject performance' (p 286). This rule of thumb likely applies to visual modelling which is perforce subject to similar shortcomings as those discussed in this section—visual stimulation of the retina is only an approximation of electrical stimulation (see [65]), plus the degenerate retina is subject to remodelling [87] and ganglion cell atrophy [88].

4. Conclusion

Further to reviewing visual modelling work pertaining to microelectronic vision prosthesis, we have sought to acquaint the vision researcher with the psychophysics and signal processing that pertains to the cochlear implant, and the use of acoustic models in that field. Acoustic modelling involves more accessible cohorts (normally hearing listeners) and allows for the separation of confounding factors (e.g., viability of the inner ear) in ongoing efforts to improve speech processors and electrode design. With the development of visual models, analogous methods and their benefits are likewise an offer to the microelectronic vision prosthesis community. In this regard, we have flagged the need for the comparison and contrast of visual modelling results with quantitative data from short-term and chronic electrical stimulation trails in blind humans (of which there is presently a paucity in the literature). There are a number of shortcomings of acoustic models that are presently being addressed; these highlight the analogous shortcomings facing the visual modelling community. Here, solutions will require not only input from human trials, but also experimental work that elucidates the electrophysiology and neurophysiology of the degenerate retina subjected to electrical stimulation.

Acknowledgments

The authors sincerely thank Professor John Morley, Dr Shaun Cloherty and Mr Yan Wong for comments on early versions of the manuscript.

References

[1] Wardrop P, Whinney D, Rebscher S J, Roland J T Jr, Luxford W and Leake P A 2005 A temporal bone study of insertion trauma and intracochlear position of cochlear implant electrodes: I. Comparison of nucleus banded and nucleus contour electrodes *Hear. Res.* **203** 54–67

[2] Rubinstein J T and Miller C A 1999 How do cochlear prostheses work? *Curr. Opin. Neurobiol.* **9** 399–404

[3] Gregory R L 1974 *Concepts and Mechanisms of Perception* (London: Duckworth)

[4] Dallos P 1992 The active cochlea *J. Neurosci.* **12** 4575–85

[5] Nobili R, Mammano F and Ashmore J 1998 How well do we understand the cochlea? *Trends Neurosci.* **21** 159–67

[6] von Helmholtz H 1885 *On the Sensations of Tone* ed A J Ellis (New York: Dover) (Engl. Transl.)

[7] Cohen N L 2004 Cochlear implant candidacy and surgical considerations *Audiol. Neuro-Otol.* **9** 197–202

[8] Gstoettner W K, Adunka O, Franz P, Hamzavi J, Plenk H Jr, Susani M, Baumgartner W and Kiefer J 2001 Perimodiolar electrodes in cochlear implant surgery *Acta Otolaryngol.* **121** 216–9

[9] Skinner M W, Holden L K, Whitford L A, Plant K L, Psarros C and Holden T A 2002 Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech processing strategies in newly implanted adults *Ear Hear.* **23** 207–23

[10] Durand D M 1995 Electrical stimulation of excitable tissue *The Biomedical Engineering Handbook* ed J D Bronzino (New York: CRC Press) pp 229–51

[11] Seligman P M and McDermott H J 1995 Architecture of the Spectra 22 speech processor *Ann. Otol. Rhinol. Laryngol.* **104** 172–5

[12] Wilson B S, Finley C F, Lawson D T, Wolford R D, Eddington D K and Rabinowitz W M 1991 Better speech recognition with cochlear implants *Nature* **352** 236–8

[13] Vandali A E, Whitford L A, Plant K L and Clark G M 2000 Speech perception as a function of electrical stimulation rate: using the new Nucleus 24 cochlear implant system *Ear Hear.* **21** 608–24

[14] Wilson B S 2004 Engineering design of cochlear implants *Cochlear Implants* ed F-G Zeng, A N Popper and R Fay (New York: Springer) pp 14–52

[15] Rubinstein J T and Hong R 2003 Signal coding in cochlear implants: exploiting stochastic effects of electrical stimulation *Ann. Otol. Rhinol. Laryngol.* **112** 14–9

[16] Friesen L M, Shannon R V and Cruz R J 2005 Effects of stimulation rate on speech recognition with cochlear implants *Audiol. Neuro-Otol.* **10** 169–84

[17] Eddington D K 1983 Speech recognition in deaf subjects with multichannel intracochlear electrodes *Ann. New York Acad. Sci.* **405** 241–58

[18] Throckmorton C S and Collins L M 2002 The effect of channel interactions on speech recognition in cochlear implant subjects: predictions from an acoustic model *J. Acoust. Soc. Am.* **112** 285–96

[19] Coleman A M 2006 *A Dictionary of Psychology* (Oxford: Oxford University Press)

[20] Blamey P J, Martin L F A and Clark G M 1984 A comparison of three speech coding strategies using an acoustic model of a cochlear implant *J. Acoust. Soc. Am.* **77** 209–17

[21] Shannon R V, Zeng F-G, Kamath V, Wygonski J and Ekelid M 1995 Speech recognition with primarily temporal cues *Science* **270** 303–4

[22] Ladefoged P 2006 *A Course in Phonetics* http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/course/contents.html (page accessed 1 Jan 2006)

[23] Miller G A and Nicely P E 1955 An analysis of perceptual confusions among some english consonants *J. Acoust. Soc. Am.* **27** 338–52

[24] Wiener N 1948 *Cybernetics* (New York: Wiley)

[25] Dorman M F, Loizou P C and Rainey D 1997 Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs *J. Acoust. Soc. Am.* **102** 2403–11

[26] Gonzalez J and Oliver J C 2005 Gender and speaker identification as a function of the number of channels in spectrally reduced speech *J. Acoust. Soc. Am.* **118** 461–70

[27] Humayun M S *et al* 2003 Visual perception in a blind subject with a chronic microelectronic retinal prosthesis *Vision Res.* **43** 2573–81

[28] Zrenner E, Besch D, Bartz-Schmidt K U, Gekeler F, Gabel V P, Kuttenkeuler C, Sachs H, Sailer H, Wilhelm B and Wilke R 2006 Subretinal chronic multi-electrode arrays implanted in blind patients *Invest. Ophthalmol. Vis. Sci.* **47** 1538 (ARVO e-abstract)

[29] Hornig R, Velikay-Parel M, Feucht M, Zehnder T and Richard G 2006 Early clinical experience with a chronic retinal implant system for artificial vision *Invest. Ophthalmol. Vis. Sci.* **47** 3216 (ARVO e-abstract)

[30] Dorman M F and Loizou P C 1998 The identification of consonants and vowels by cochlear implant patients using a

6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels *Ear Hear.* **19** 162–6
[31] Dorman M F, Loizou P C and Fitzke J 1998 The identification of speech in noise by cochlear implant patients and normal-hearing listeners using six-channel signal processors *Ear Hear.* **19** 481–4
[32] Friesen L M, Shannon R V, Baskent D and Wang X 2001 Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants *J. Acoust. Soc. Am.* **110** 1150–63
[33] Fishman K E, Shannon R V and Slattery W H 1997 Speech recognition as a function of the number of electrodes used in the speak cochlear implant speech processor *J. Speech Language Hear. Res.* **40** 1201–15
[34] Merritt C D and Justus B L 2003 Fabrication of microelectrode arrays having high-aspect-ratio microwires *Chem. Mater.* **15** 2520–6
[35] Hallum L E, Chen S C, Cloherty S L and Lovell N H 2006 Psychophysics of prosthetic vision: II. Stochastic sampling, the phosphene image, and noise *Proc. 28th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*
[36] Dorman M F, Loizou P C and Rainey D 1997 Simulating the effect of cochlear-implant electrode insertion depth on speech understanding *J. Acoust. Soc. Am.* **102** 2993–6
[37] Greenwood D D 1990 A cochlear frequency-position function for several species—29 years later *J. Acoust. Soc. Am.* **87** 2592–605
[38] Baskent D and Shannon R V 2003 Speech recognition under conditions of frequency–place compression and expansion *J. Acoust. Soc. Am.* **113** 2064–76
[39] Xu L, Thompson C S and Pfingst B E 2005 Relative contribution of spectral and temporal cues for phoneme recognition *J. Acoust. Soc. Am.* **117** 3255–67
[40] Fu Q-J and Nogaki G 2004 Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing *J. Assoc. Res. Otolaryngol.* **6** 19–27
[41] Loizou P C, Dorman M, Poroy O and Spahr T 2000 Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution *J. Acoust. Soc. Am.* **108** 2377–87
[42] Drullman R 1995 Temporal envelope and fine structure cues for speech intelligibility *J. Acoust. Soc. Am.* **97** 585–92
[43] Weiland J D, Liu W T and Humayun M S 2005 Retinal prosthesis *Annu. Rev. Biomed. Eng.* **7** 361–401
[44] Zrenner E 2002 Will retinal implants restore vision? *Science* **295** 1022–5
[45] Suaning G J, Hallum L E, Chen S C, Preston P J and Lovell N H 2003 Phosphene vision: development of a portable visual prosthesis system for the blind *Proc. 25th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)* pp 2047–50
[46] Chow A Y, Pardue M T, Perlman J I, Ball S L, Chow V Y, Hetling J R, Peyman G A, Liang C P, Stubbs E B and Peachey N S 2002 Subretinal implantation of semiconductor-based photodiodes: durability of novel implant designs *J. Rehabil. Res. Dev.* **39** 313–21
[47] Humayun M S, de Juan E Jr, Dagnelie G, Greenberg R J, Propst R H and Phillips D H 1996 Visual perception elicited by electrical stimulation of the retina in blind humans *Arch. Ophthalmol.* **114** 40–6
[48] Weiland J D, Humayun M S, Dagnelie G, de Juan E, Greenberg R J and Iliff N T 1999 Understanding the origin of visual percepts elicited by electrical stimulation of the human retina *Graef's Arch. Clin. Exp. Ophthalmol.* **237** 1007–13
[49] Rizzo J F III, Wyatt J, Loewenstein J, Kelly S and Shire D 2003 Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials *Invest. Ophthalmol. Vis. Sci.* **44** 5362–9
[50] Richard G, Feucht M, Bornfeld N, Laube T, Rossler G, Velikay-Parel M and Hornig R 2005 Multicenter study on acute electrical stimulation of the human retina with an epiretinal implant: clinical results in 20 patients *Invest. Ophthalmol. Vis. Sci.* **46** (Suppl. S) 1143
[51] Cha K, Horch K W, Normann R A and Boman D K 1992 Reading speed with a pixelized vision system *J. Opt. Soc. Am. A* **9** 673–7
[52] Cha K, Horch K W and Normann R A 1992 Simulation of a phosphene-based visual field: visual acuity in a pixelized vision system *Ann. Biomed. Eng.* **20** 439–49
[53] Cha K, Horch K W and Normann R A 1992 Mobility performance with a pixelized vision system *Vision Res.* **32** 1367–72
[54] Humayun M S 2001 Intraocular retinal prosthesis *Trans. Am. Ophthalmol. Soc.* **99** 271–300
[55] Sommerhalder J, Oueghlani E, Bagnoud M, Leonards U, Safran A B and Pelizzone M 2003 Simulation of artificial vision: I. Eccentric reading of isolated words, and perceptual learning *Vision Res.* **43** 269–83
[56] Sommerhalder J, Rappaz B, de Haller R, Pérez Fornos A, Safran A B and Pelizzone M 2004 Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task *Vision Res.* **44** 1693–706
[57] Pérez Fornos A, Sommerhalder J, Rappaz B, Safran A B and Pelizzone M 2005 Simulation of artificial vision: III. Do the spatial or temporal characteristics of stimulus pixelization really matter? *Invest. Ophthalmol. Vis. Sci.* **46** 3906–12
[58] Thompson R W Jr, Barnett G D, Humayun M S and Dagnelie G 2003 Facial recognition using simulated prosthetic pixelized vision *Invest. Ophthalmol. Vis. Sci.* **44** 5035–42
[59] Hayes J S, Yin V T, Piyathaisere D, Weiland J D, Humayun M S and Dagnelie G 2003 Visually guided performance of simple tasks using simulated prosthetic vision *Artif. Organs* **27** 1016–28
[60] Chen S C, Hallum L E, Lovell N H and Suaning G J 2005 Visual acuity measurement of prosthetic vision: a virtual-reality simulation study *J. Neural Eng.* **2** S135–45
[61] Chen S C, Hallum L E, Lovell N H and Suaning G J 2005 Learning prosthetic vision: a virtual-reality study *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 249–55
[62] Hallum L E, Suaning G J, Taubman D S and Lovell N H 2005 Simulated prosthetic visual fixation, saccade, and smooth pursuit *Vision Res.* **45** 775–88
[63] Dagnelie G, Barnett D, Humayun M S and Thompson R W 2006 Paragraph text reading using a pixelized prosthetic vision simulator: parameter dependence and task learning in free-viewing conditions *Invest. Ophthalmol. Vis. Sci.* **47** 1241–50
[64] Dagnelie G 2006 Playing checkers: detection and eye–hand coordination in simulated prosthetic vision *J. Mod. Opt.* **53** 1325–42
[65] Cottaris N P and Elfar S D 2005 How the retinal network reacts to epiretinal stimulation to form the prosthetic visual input to the cortex *J. Neural Eng.* **2** S74–90
[66] Hancock P *Psychological Image Collection at Stirling* http://pics.psych.stir.ac.uk (pages accessed 1 July 2003)
[67] Hallum L E, Chen S C, Preston P J, Suaning G J and Lovell N H 2005 Simulating prosthetic vision *Invest. Ophthalmol. Vis. Sci.* **46** 1522 (ARVO e-abstract)
[68] Boyle J R 2005 Improving perception from electronic visual prostheses *PhD Thesis* Queensland University of Technology
[69] Suaning G J and Lovell N H 2001 CMOS neurostimulation ASIC with 100 channels, scaleable output, and bi-directional radio-frequency telemetry *IEEE Trans. Biomed. Eng.* **48** 248–60

[70] Sivaprakasam M, Liu W, Humayun M S and Weiland J D 2005 A variable range bi-phasic current stimulus driver circuitry for an implantable retinal prosthetic device *IEEE J. Solid-State Circuits* **40** 763–71
[71] Pelli D G 1987 The visual requirements of mobility *Low Vision: Principles and Applications* ed G C Woo (New York: Springer) pp 134–45
[72] Harmon L D 1973 The recognition of faces *Sci. Am.* **229** 70–83
[73] House W 1985 A personal perspective on cochlear implants *Cochlear Implants* ed R Schindler and M Merzenich (New York: Raven Press) pp 13–6
[74] Loizou P C 1998 Mimicking the human ear *IEEE Signal Process. Mag.* **15** 102–30
[75] Hallum L E, Suaning G J and Lovell N H 2004 Contribution to the theory of prosthetic vision *ASAIO J.* **50** 392–6
[76] Fu L, Cai S, Zhang H, Hu G and Zhang X 2006 Psychophysics of reading with a limited number of pixels: towards the rehabilitation of reading ability with visual prosthesis *Vision Res.* **46** 1292–301
[77] Legge G E, Pelli D G, Rubin G S and Schleske M M 1985 Psychophysics of reading: I. Normal vision *Vision Res.* **25** 239–52
[78] Dokos S, Suaning G J and Lovell N H 2005 A bidomain model of epiretinal stimulation *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 137–46
[79] Hallum L E, Suaning G J, Taubman D S and Lovell N H 2004 Towards photosensor movement-adaptive image analysis in an electronic retinal prosthesis *Proc. 26th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)* pp 4165–8
[80] Morrone M C, Burr D C and Ross J 1983 Added noise restores recognizability of coarse quantized images *Nature* **305** 226–8
[81] Harmon L D and Julesz B 1973 Masking in visual recognition—effects of 2-dimensional filtered noise *Science* **180** 1194–7
[82] Hallum L E, Cloherty S L, Taubman D S, Suaning G J and Lovell N H 2006 Psychophysics of prosthetic vision: III. Stochastic rendering, the phosphene image, and perception *Proc. 28th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*
[83] Leigh R J and Zee D S 1999 *The Neurology of Eye Movements* 3rd edn (New York: Oxford University Press)
[84] Shepherd R K and Hardie N A 2001 Deafness-induced changes in the auditory pathway: implications for cochlear implants *Audiol. Neuro-Otol.* **6** 305–18
[85] Kiang N Y S and Moxon E C 1972 Physiological considerations in artificial stimulation of the inner ear *Ann. Otol. Rhinol. Laryngol.* **81** 714–30
[86] Rubinstein J T, Wilson B S, Finley C C and Abbas P J 1999 Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation *Hear. Res.* **127** 108–18
[87] Jones B W and Marc R E 2005 Retinal remodeling during retinal degeneration *Exp. Eye Res.* **81** 123–37
[88] Stone J L, Barlow W E, Humayun M S, de Juan E and Milam A H 1992 Morphometric analysis of macular photoreceptors and ganglion-cells in retinas with retinitis-pigmentosa *Arch. Ophthalmol.* **110** 1634–9