Hearing Research 456 (2025) 109163

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Ensemble responses of auditory midbrain neurons in the cat to speech stimuli at different signal-to-noise ratios

Anu Sabu [a,c,*], Dexter Irvine [a,b], David B. Grayden [a,d], James Fallon [a,c]

[a] Bionics Institute, Fitzroy, Victoria, Australia
[b] School of Psychological Sciences, Monash University, Clayton, Victoria, Australia
[c] Medical Bionics Department, The University of Melbourne, Parkville, Victoria, Australia
[d] Department of Biomedical Engineering and Graeme Clark Institute, The University of Melbourne, Melbourne, Victoria, Australia

ARTICLE INFO

Keywords:
Speech discrimination in noise
Neural discrimination
Acoustic stimulation
Signal-tonoise ratios
Inferior colliculus recordings

ABSTRACT

Originally reserved for those who are profoundly deaf, cochlear implantation is now common for people with partial hearing loss, particularly when combined with a hearing aid. This combined intervention enhances speech comprehension and sound quality when compared to electrical stimulation alone, particularly in noisy environments, but the physiological basis for the benefits is not well understood. Our long-term aim is to elucidate the underlying physiological mechanisms of this improvement, and as a first step in this process, we have investigated in normal hearing cats, the degree to which the patterns of neural activity evoked in the inferior colliculus (IC) by speech sounds in various levels of noise allows discrimination between those sounds. Neuronal responses were recorded simultaneously from 32 sites across the tonotopic axis of the IC in anaesthetised normal hearing cats ($n = 7$). Speech sounds were presented at 20, 40 and 60 dB SPL in quiet and with increasing levels of additive noise (signal-to-noise ratios (SNRs) –20, –15, –10, –5, 0, +5, +10, +15, +20 dB). Neural discrimination was assessed using a Euclidean measure of distance between neural responses, resulting in a function reflecting speech sound differentiation across various SNRs. Responses of IC neurons reliably encoded the speech stimuli when presented in quiet, with optimal performance when an analysis bin-width of 5–10 ms was used. Discrimination thresholds did not depend on stimulus level and were best for shorter analysis binwidths. This study sheds light on how the auditory midbrain represents speech sounds and provides baseline data with which responses to electro-acoustic speech sounds in partially deafened animals can be compared.

1. Introduction

The expansion of criteria for cochlear implantation to include people with residual low-frequency hearing, and the development of minimally traumatic implantation procedures to preserve this hearing, has focused interest on the benefits of combined electroacoustic stimulation (EAS). In EAS, electrical stimulation is delivered to the basal, high-frequency end of the cochlea while acoustic stimulation is delivered to the apical, low-frequency end (typically with acoustic amplification over the range of residual hearing). There is consensus that EAS provides better speech perception, particularly of speech in noise, than electrical stimulation alone (Cantore et al., 2020; Gantz et al., 2018; Gifford et al., 2018; Incerti et al., 2013; Wilson, 2010). However, as noted by Sato et al. (2017), the physiological bases for the benefits provided by EAS, and the interactions between the activity evoked by acoustic and electrical stimulation in the central auditory system, are not well understood.

To the best of our knowledge, there are no studies that report the encoding of combined electrical and acoustic stimulation to complex stimuli within the central auditory pathway. The only accounts of central auditory system responses to combined electric and acoustic stimulation are those of Vollmer et al. (2010) and Noh and Lee (2012), in normal hearing cats and guinea pigs, respectively. Both studies employed simple electric (short pulse trains or sinusoids) and acoustic (pure tones or noise) stimuli and investigated masking between the two stimulus types in the inferior colliculus (IC). Understanding the processes underlying the clinically observed enhancement of speech perception requires studies that use more complex sounds, such as speech sounds, and the electrical stimuli generated by an EAS-programmed cochlear implant processing those sounds. This is

* Corresponding author.
E-mail address: asabu@bionicsinstitute.org (A. Sabu).

https://doi.org/10.1016/j.heares.2024.109163
Received 17 September 2024; Received in revised form 13 November 2024; Accepted 2 December 2024
Available online 3 December 2024

particularly important as clinical cochlear implants necessarily introduce a processing delay relative to direct acoustic stimulation and broad regions of synchronised cochlea activation, both of which may be expected to degrade information transfer.

To elucidate the neural processing of acoustically and electrically presented speech in the central auditory pathway of listeners with severe deafness, we plan to investigate the representation of EAS speech stimuli in the midbrain of a neonatally partially deafened, implanted cat model (Irving et al., 2014). As the first step in this research program, and to obtain baseline data against which the data from partially deafened cats can be compared, we have studied the ensemble representation of simple speech sounds (consonant-vowel-consonant (CvC) stimuli) at different signal-to-noise ratios (SNRs) by simultaneous recordings from multiple sites across the tonotopic axis of the central nucleus of IC (CIC) in normal-hearing cats.

There have been numerous studies of the coding of speech (and other vocalizations) in the inferior colliculus (see Portfors and Sinex (2005) for a review). Human studies have generally recorded evoked potentials (Chandrasekaran et al., 2012; Johnson et al., 2008; Presacco et al., 2019), and animal studies have recorded single or multiunit activity from a single site within the nucleus (Sinex and Chen, 2000; Warrier et al., 2011). In contrast, Perez et al. (2013) recorded single and multiunit activity from a number of sites in the IC of each of 27 anesthetized rats. The fundamental frequencies and envelopes of their CvC sounds were shifted up in frequency by one octave in order to better match the rat hearing range. They generated neurograms of the population responses to each sound by combining (across animals) the responses of all 187 recording sites, ordered by the characteristic frequency (CF) of the neurons at each site. They also trained rats on behavioural vowel and consonant discrimination tasks and compared neural and behavioural discrimination. They reported that performance on vowel discrimination tasks was highly correlated with neural discrimination based on spike count but was not correlated when spike timing information was preserved. In contrast, consonant discrimination performance was highly correlated with neural discrimination when spike timing was preserved, but not when spike timing was eliminated.

Our data extends these previous studies by recording sufficient multiunit activity across the IC from each animal to allow neural discrimination to be determined for individual animals, allowing an assessment of normal variation. Our data also allows us to examine the hypothesis that at the midbrain level, vowel sounds are encoded in terms of spike count and consonant sounds in terms of spike timing.

2. Materials and methods

2.1. Experimental subjects

The basic surgical and electrophysiological procedures used to obtain recordings from normal-hearing adult cats ($n = 7$) have been described in detail previously (George et al., 2015) and so will be described only briefly in the following sections. All procedures were in accordance with Australian Code of Practice for the Care and Use of Animals for Scientific Purposes and with the Guidelines laid down by the National Institutes of Health in the USA regarding the care and use of animals for experimental procedures and were approved by the St Vincent's Hospital Animal Research and Ethics Committee.

2.2. Anaesthesia and surgery

Animals were anesthetized using a mixture of methadone (0.4 mg/kg, subcutaneous), medetomidine (12μg/kg, subcutaneous) and ketamine (8mg/kg, subcutaneous), and maintained over the duration of the experiment (3 days) with a constant rate infusion through intravenous supply of propofol 5–12 mg/kg hr^{-1}) and methadone (0.01 mg/mL) in the right and/or left cephalic vein to effect. Ventilation was maintained with 100 % O_2 using an endotracheal tube. Fluids were maintained via delivery in Hartmann's (3 mL/kg hr^{-1}, intravenous) and physiological parameters (heart rate, respiratory rate, SpO_2 %, $EtCO_2$, NIBP, and temperature) were maintained at normal physiological levels. Assessment of corneal and pedal reflexes were made every 30–60 mins to ensure adequate depth of anaesthesia was maintained.

Animals were placed in a modified stereotaxic frame that left the ears free of obstruction. A craniotomy was performed, and the cerebral cortex was aspirated to reveal the dorsal surface of the right IC. If required, a portion of the tentorium was removed using a small diamond burr to expose the entire dorsolateral surface of the IC.

2.3. Normal hearing screening

To screen for normal hearing, tone-evoked auditory brain stem responses (ABRs) were recorded using previous published techniques (Irving et al., 2014). Briefly, subcutaneous electrodes were placed at vertex (positive), nape of the neck (negative) and thorax (ground). Signals were differentially amplified $x10^3$ via an ISO-80 Bio-Amplifier (World Precision Instruments, Sarasota, FL) and bandpass filtered between 150 Hz and 3 kHz. Responses from 100 repetitions of each tone-pip (5ms duration, 1 ms linear rise/fall, 25 stimuli/s; 0.5, 1, 2, 4, 8, 12, 24 and 36 kHz presented at 30 - 100 dB SPL in 5dB-SPL steps) were averaged and displayed using IgorPro (Wavemetrics, USA). Threshold was defined as a response of at least 0.1 μV in two consecutive responses. Only animals whose thresholds were in the normal-hearing range were used in this study.

2.4. Recording procedures

All recordings were made in a double-walled sound attenuating chamber. Multi-unit activity was recorded simultaneously from all the channels on an array (single shank, 32 channel array, pitch 100 μm, NeuroNexus, Ann, Arbor, MI) inserted across the tonotopic axis of the CIC (Fig. 1a). The array recorded neural activity over approximately 3.1 mm of the IC covering a range of characteristic frequencies from 500 Hz to 30kHz. Neural spike activity from the 32 recording sites was amplified, band-pass-filtered (0.3–5 kHz) and digitized at a sample rate of 30 kHz using a Cerebus data acquisition system (Blackrock Microsystems, USA). For each recording site, the response to tone-pips (5-ms rise/fall, 100 ms duration) at various frequency-intensity combination was plotted to produce a characteristic response area image (Fig. 1b). These response area images were used to ensure the array was position such that recordings were obtained across the entire tonotopic axis of the CIC.

2.5. Stimuli and data analysis

All stimuli were presented 'free field' via a calibrated speaker positioned 10 cm from the pinna contralateral to the recording site, at rates of 0.5 to 1.0 stimuli/*sec*. To assess the neural responses to simple speech tokens a series of CvC tokens were presented at 20, 40 and 60 dB SPL in both quiet, and in various levels of background noise (white noise presented at signal-to-noise ratios (SNRs):20, –15, –10, –5, 0, +5, +10, +15, +20 dB SPL). The background noise level in the sound-attenuating chamber (i.e., the noise level in the "quiet" condition) was below 20 dB SPL. CvC tokens consisted of BvDs (bad, bed, bid, bod, bud, bead, board, bird, bard, bared, booed), HvPs (hap, hep, hip, hop, hup, heap, hoarp, hirp, harp, harep, hooep) and HvDs (had, hed, hid, hod, hud, head, hoard, hird, hard, hared, hooed). The CvC tokens used in this study were recordings of a female voice with fundamental frequency around 170 Hz and first formant frequency around 340Hz.

The recorded neural activity was processed offline using customized spike detection scripts in IgorPro. Stimulus artefacts were removed (Heffer and Fallon, 2008) and spike detection thresholding (all events that crossed −4 RMS of baseline, resulting in an effective minimum retrigger time of 3 samples (90μs)) was applied (Fallon et al., 2009) to detect the multi-unit spiking activity in the 5-45 ms post stimulus

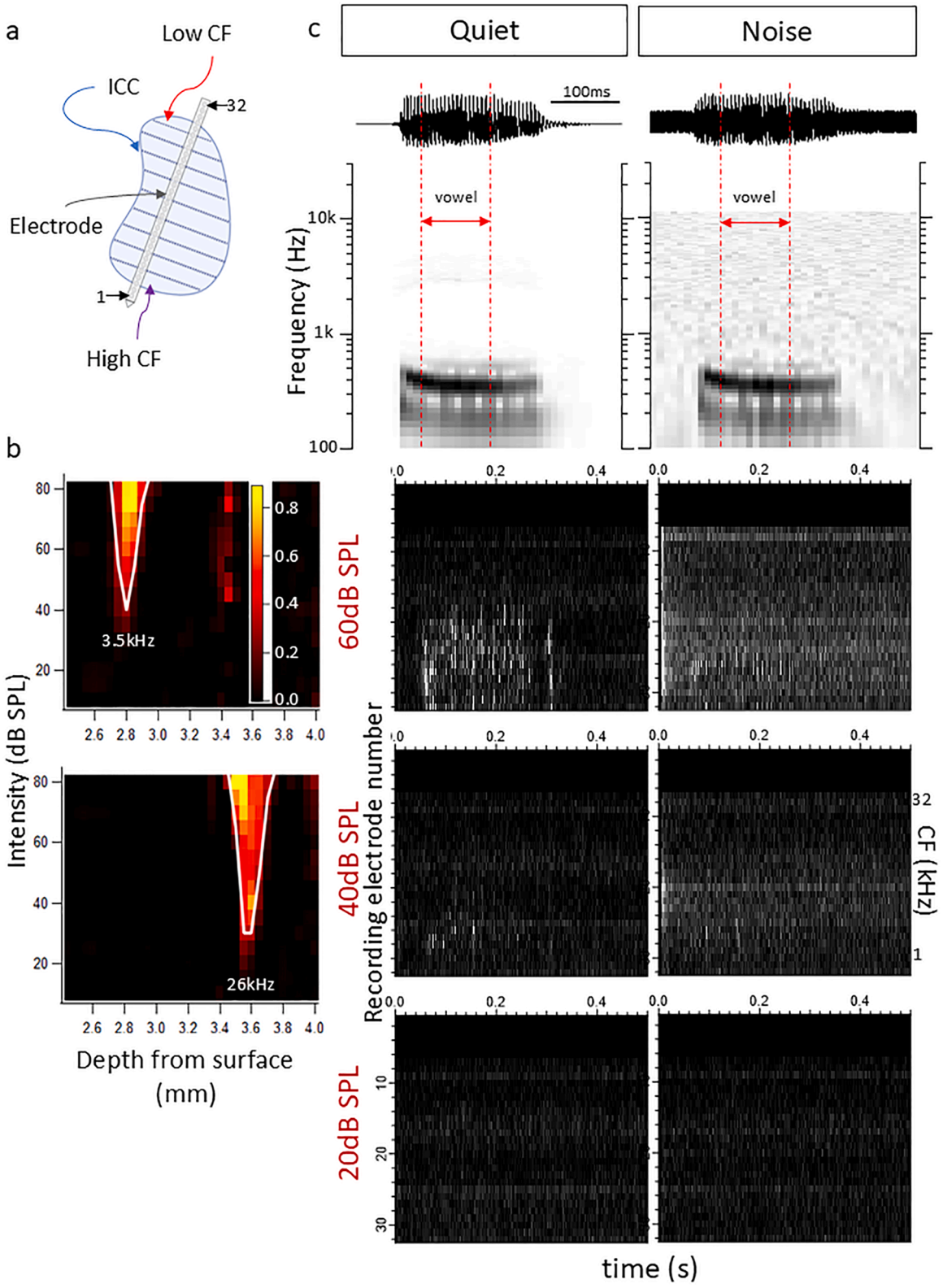

Fig. 1. (a) Schematic diagram showing a multichannel recording array oriented along the tonotopic axis of the central nucleus of the inferior colliculus (CIC). The recording electrode number 1 records high frequency response while recording electrode number 32 records low frequency response. (b) Sample acoustic response images to 3.5 kHz and 26 kHz pure tones. (c) Sample spectrograms recorded from the CIC in cat 1806 in response to the word 'bead' processed at a 10 ms bin width and sample neurograms processed at a 1 ms bin width presented at three different levels (20, 40, and 60 dB SPL) in quiet and in noise, at a signal-to-noise ratio (Noise) of +10. For the neurograms, the left axis shows the recording electrode number, and the right axis shows the approximate characteristic frequency in kHz.

window for response area analysis (George et al., 2015).

The responses to the CvC tokens were analysed by constructing 'neurograms' from the multi-unit spiking activity, in which the response is represented as a heat-map of activity plotted across the recording array location (vertical axis) and time (binned along the horizontal axis). Examples of the neurograms for 'bead' at 20-, 40-, and 60-dB SPL in quiet and at +10 dB SNR are shown in Fig. 1c. Each neurogram was converted to an n-dimensional representation, where the number of dimensions was determined by the number of recording electrodes and response bins. For each animal, exemplar responses for each speech token in quiet were then computed as the mean (n = 60–120 stimulus presentations) for that token. The minimum Euclidean distance between the neurogram for an individual stimulus and the exemplar responses was then used to classify each response. Neural discrimination (d') was computed from the hit-rate and false-alarm rate for these classifications and the neural discrimination threshold was defined as the lowest SNR at which d' crossed 1 and, therefore, was limited to be 20 dB or less.

The analysis window of 400 ms was used for the whole CvC analysis. For vowel-only analysis, a segment of recording from 100 ms to 200 ms was used, excluding the initial consonant and final consonant in each speech token.

The CvC classification analysis was performed in Igor using a custom script. All statistical analyses were performed using the IBM SPSS statistics software. Repeated measures of analysis of variance (RM ANOVA) were carried out to determine the effect of bin width on neural discrimination and also the effect of the vowel-only component versus the whole CvC on the discrimination threshold in quiet and varying levels of background noise.

3. Results

As indicated by the representative responses in Fig. 1c (from cat 1806), responses to the speech sounds were more clearly visible at 60 dB SPL than at 20- and 40-dB SPL. The onset and offset of the speech token ('bead') are clearly visible in the 60 dB quiet neurogram as broad bands of activation across a large portion of the recording array at ~20 and ~320 ms, respectively. The vowel component is visible as bands of activity repeating at approximately 180 Hz from 30 to 230ms. The addition of noise to create an SNR of +10 dB resulted in broad onset response across all active channels on the array at 60 dB SPL and sustained

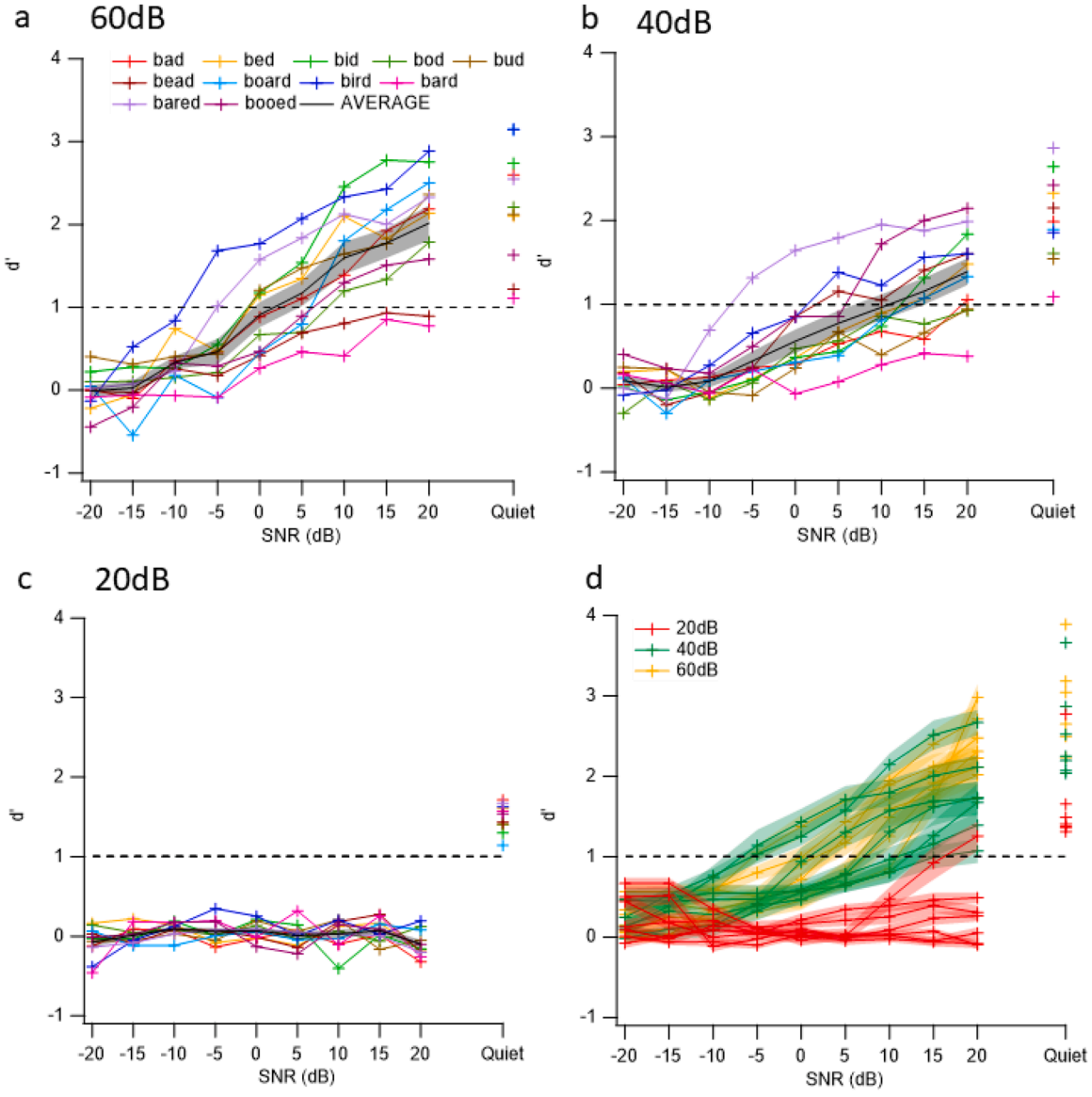

Fig. 2. Responses for different BvDs processed with a 1 ms bin width for one animal (#1806) for stimuli presented at intensities a) 60 dB, b) 40 dB and c)20 dB d) The mean detection thresholds for each of the animals (n = 7) averaged over BvDs. Shading represents the standard error of mean for the 11 BvD tokens.

activity across the duration of the noise presentation, reducing the contrast of the speech-induced activity.

Fig. 2a-c illustrates the effects of changing SNR on the neural discrimination of BvDs over a range of stimulus intensities for this animal, utilising a 1-ms analysis bin width. At a presentation level of 60 dB SPL (Fig. 2a), all BvDs were discriminable (d' > 1) in quiet. As the background noise level was increased, the discriminability decreased (d' reduced), such that the average discrimination threshold (d' = 1) was 5 ± 3 dB Results for stimuli presented at 40 dB SPL (Fig. 2.b) were broadly similar, with an average discrimination threshold of 12 ± 3 dB In contrast, at 20 dB SPL (Fig. 2.c), only stimuli presented in quiet were discriminable, with the addition of all levels of noise resulting in d' of approximately 0. The average neural discrimination for each animal (n = 7) are illustrated in Fig 2.d. It was possible in only one animal (#1813) to determine a discrimination threshold for 20 dB SPL stimuli. The average discrimination threshold across all animals for 40 dB SPL stimuli was 7 ± 3 dB which was not significantly different (paired t-test, p = 0.204) to the 4 ± 2 dB discrimination threshold for 60 dB SPL stimuli.

In an attempt to assess the influence of temporal information on neural discrimination, we repeated the analysis using bin widths ranging from 0.1 to 400ms. Data for cat #1806 are presented in Fig. 3. At a presentation level of 60 dB SPL (Fig. 3.a), increasing the bin width resulted in decreased discrimination. The reduction in discrimination is presumably a consequence of the decrease in the amount of temporal information captured by the analysis. Notably, if there is a single analysis bin (400 ms) that extends over the entire duration of the BvD, d' drops below 1 even in quiet. Results for stimuli presented at 40 dB SPL are broadly similar (Fig. 3.b), although discrimination with smaller bin widths (e.g. < 1 ms) is also degraded. The reduction in discrimination is likely driven by the low spikes counts that result from short windows, resulting in additional quantisation of the response and an increase in the number of 'empty' bins. In particular, given the average width of a single spike in the multi-unit data was ~300μs, for the 0.1 ms bin width spikes counts were quantised to 0 or 1. No effect of bin width on discrimination threshold is measurable for 20 dB SPL stimuli (Fig. 3.c).

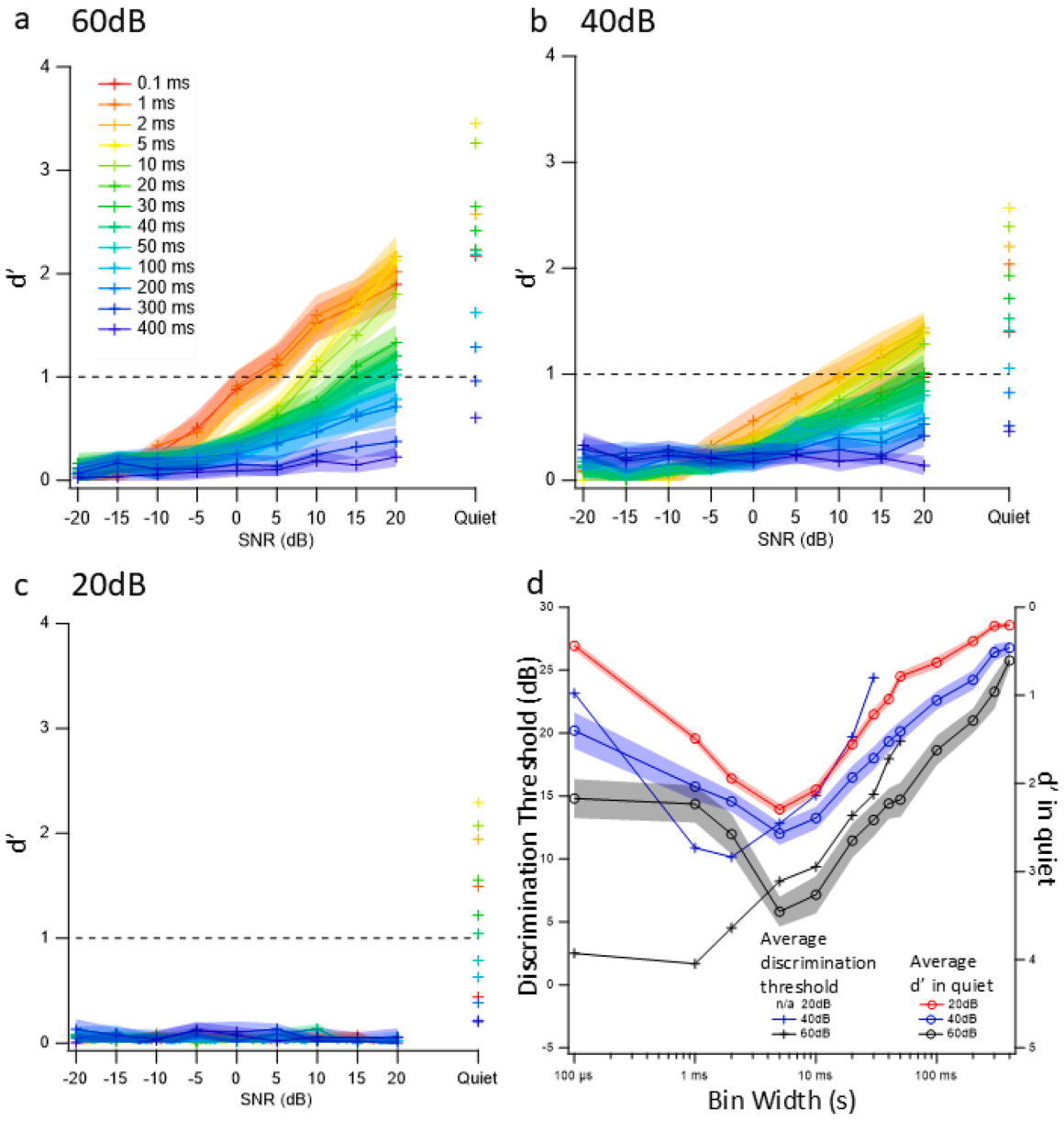

Fig. 3. Average response of all BvD CvCs processed at bin widths of 0.1 to 400 ms, for one animal (#1806) for stimuli presented at a)60 dB, b)40 dB and c)20 dB d) The SNR discrimination threshold and the d' in quiet for #1806 plotted across the different bin widths for 60 dB (black), 40 dB (blue) and 20 dB (red) SPL.

In contrast to the largely monotonic improvement in discrimination in noise with decreasing bin width, there was a non-monotonic effect of bin width on d' in quiet (Fig. 3.d). Specifically, there is an 'optimal' bin width in the 5–10 ms range which results in the maximum d' in quiet. This optimal bin width for d' in quiet was broadly similar across stimulus level. Utilising bin widths of 20 ms or longer resulted in similar decreases in d' across all stimulation levels. Similarly, bin widths of 2 ms or shorter resulted in degradation of discrimination.

The interplay between stimulus intensity and analysis bin width on d' in quiet described in Fig. 3 for an individual animal for BvD stimuli was robust across both animals and speech tokens (Fig. 4). Maximum d' was achieved with a bin width of 5–10 ms at all levels and with all three sets of CvCs. A two-way (Level x CvC type) RM ANOVA on d' with a 5 ms bin width showed there was a significant effect of Level ($p < 0.001$, Bonferroni post hoc test: p(20 dB, 40 dB)=0.016, p(40 dB, 60 dB)= 0.005, p(20 dB, 60 dB)<0.001) but no effect of CvC type ($p = 0.241$) or interaction ($p = 0.681$).

As noted above, the discrimination threshold for 20 dB SPL sounds could not be measured in most animals (6 of 7), as it was above the maximum SNR tested (20 dB), so was not further analysed. Discrimination threshold was lowest for shorter bin widths (Fig. 5), with a 1 ms bin width producing the minimum discrimination thresholds. A two-way (Level x CvC type) RM ANOVA on discrimination threshold with a 1 ms bin width showed there was no significant effect of Level ($p = 0.873$) or CvC type ($p = 0.294$) or interaction ($p = 0.874$).

The above results highlight the importance of spike timing information for encoding CvCs in both quiet and noisy conditions. Perez et al. (2013) demonstrated that spike timing was important for consonant, but not vowel discrimination, so we reanalysed our data and focused on the vowel portion of the CvC tokens (Fig. 6). For 40 dB SPL stimuli, there was a significant difference in discrimination threshold (Fig. 6.a blue) when analysis was restricted to the vowel component compared to the whole CvC (2-way RM ANOVA, $p = 0.012$). However, there was no significant effect of the bin width ($p = 0.083$) on the discrimination threshold in noise and no significant interaction ($p = 1$) between the types (vowel-only component, whole CvC) and the bin width, for the 40 dB stimuli presentations. In contrast, for 60 dB SPL stimuli, there was a significant increase in discrimination threshold (Fig. 6.a black) when analysis was restricted to the vowel component (2-way RM ANOVA, $p < 0.001$), and with increasing bin widths ($p < 0.001$), and no interaction ($p = 0.994$) between the two. A post hoc analysis of the discrimination threshold with Bonferroni correction showed that the discrimination threshold was significantly elevated for bin widths of 20 ms and greater, regardless of the type of input; i.e., the vowel-only component or the whole CvC. In noise, the performance deteriorated with increasing bin widths for the 60 dB stimuli presentations. The minimum discrimination threshold occurring for the smaller bin width is consistent with spike timing information being important for encoding vowels.

The d' in quiet was lower for vowel-only analysis compared to the whole CvC analysis for all conditions tested (Fig. 6b). Statistical analysis of the d' data (2-way RM ANOVA (Type, Bin Width)) indicated significant main effects and interactions at all 3 stimulus levels (all $p < 0.001$). There was no statistically significant interaction between the Type and Bin Width for 40 dB ($p = 0.983$) or 60 dB ($p = 0.898$) SPL stimuli, but there was for 20 dB ($p = 0.037$) SPL stimuli. There was a statistically significant interaction between the Level and Type ($p < 0.001$). Of note, the maximum d' was achieved with 5–10 ms bin width for all analyses apart from vowel-only at 20 dB SPL. For the vowel-only at 20 dB SPL condition, there was a clear shift for maximum d' to occur at shorter (1–5 ms) bin widths, suggesting timing information could be particularly important for encoding vowels in quiet conditions. However, it is worth noting that the smallest bin width of 0.1 ms never resulted in the maximum d', suggesting there is a limit to the information being encoded.

All results presented thus far have utilised >100 repetitions of each stimulus-noise combination, requiring substantial (>48hours) recording time. In an effort to determine the minimum number of repetitions required, we repeated the above analysis utilising the near optimal 10 ms bin width on subsets of the data to determine the effect of the number of repetitions on the analysis. Fig. 7 illustrates that there is no effect of halving the number of repetitions (from 120 to 60) (2-way RM ANOVA, $p = 0.819$), for 40 and 60 dB SPL ($p = 0.666$) stimuli (data gathered around line of proportionality) and no interaction ($p = 0.998$) between the stimulus level and number of repetitions on the discrimination threshold.

To examine the errors in discrimination, a confusion matrix based on data from all the animals was determined for the different signals presented at 60, 40 and 20 dB SPL (Fig. 8). As smaller bin widths gave better performance, data were processed at a bin width of 10 ms in quiet and at +10 dB SNR. In quiet, with a 10 ms bin width analysis, there was little

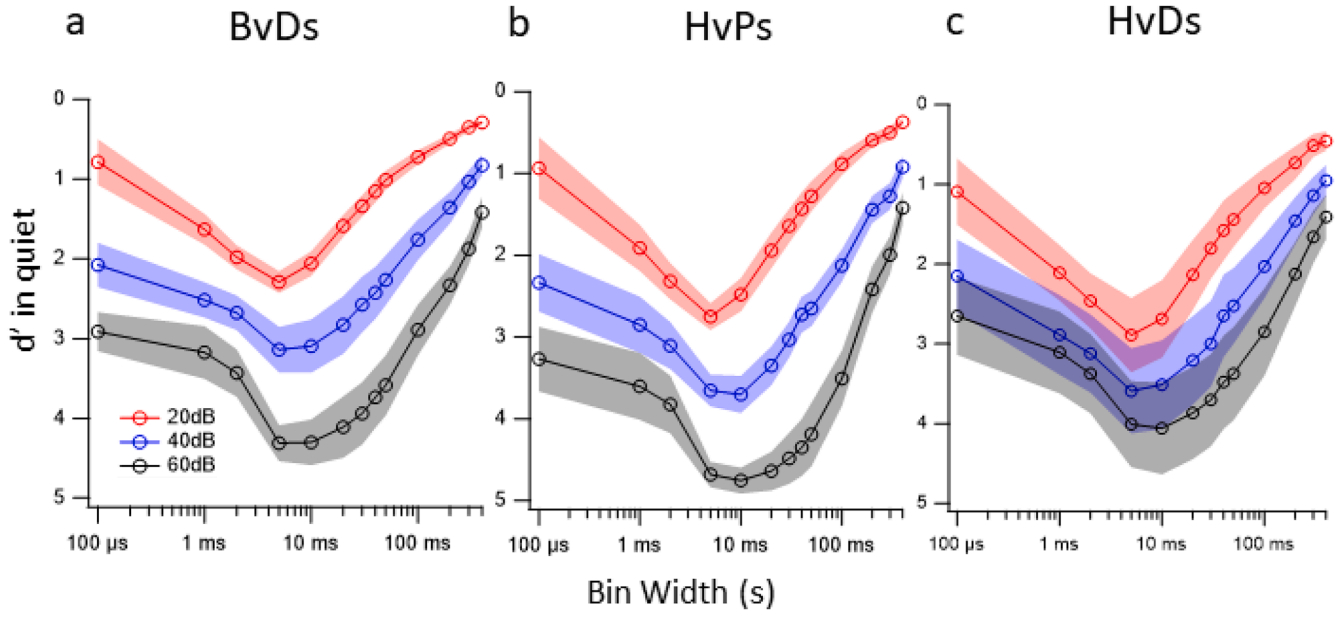

Fig. 4. Average response of all CvCs a)BvDs, b)HvPs and c)HvDs processed at bin widths of 0.1 to 400 ms, for all animals ($n = 7$) for stimuli presented at 60 dB (black), 40 dB (blue) and 20 dB (red). The shading represents the standard error of the mean across the 7 animals.

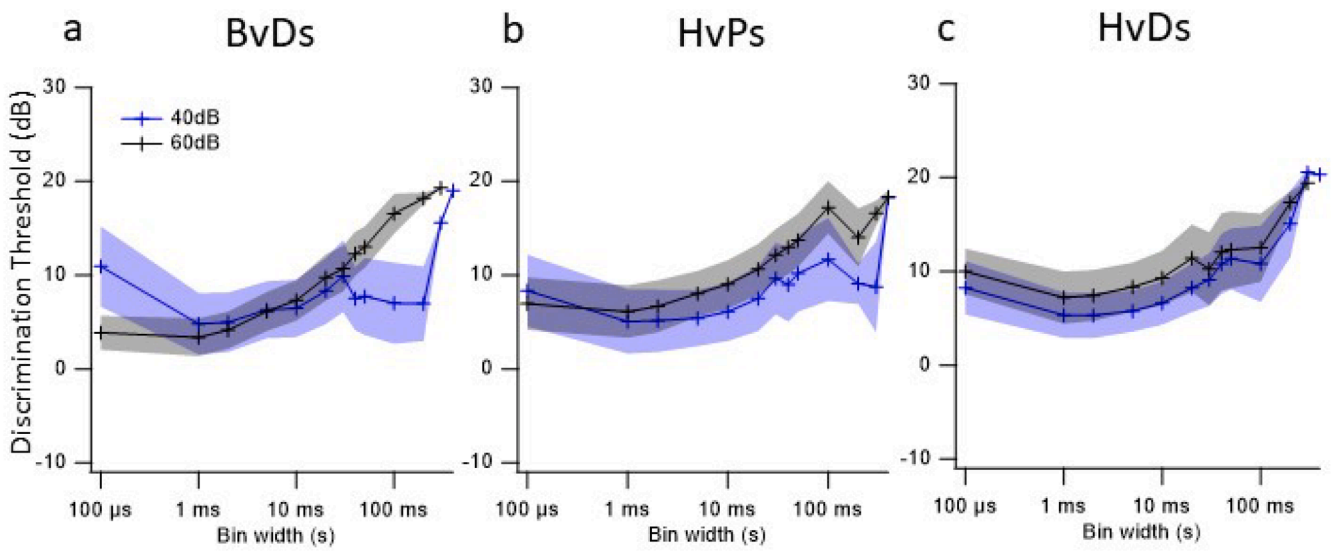

Fig. 5. Average discrimination threshold for all CvCs a) BvDs, b) HvPs and c) HvDs processed at bin widths of 0.1 to 400 ms, for all animals ($n = 7$) for stimuli presented at 40 dB (blue) and 60 dB (black). The error bars represent the standard error of mean across the 7 animals.

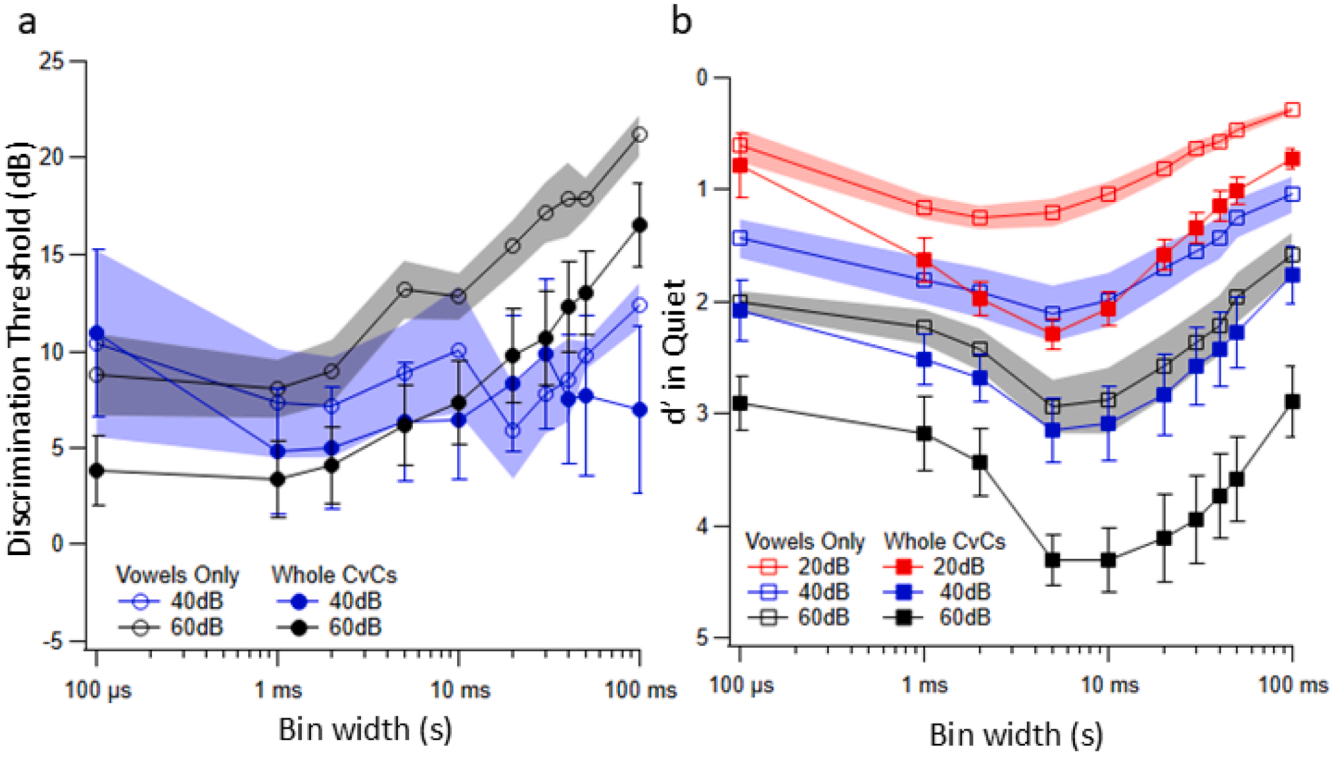

Fig. 6. a) Average discrimination threshold for vowels-only and whole token analysis of the BvD tokens processed for bin widths of 0.1 to 100 ms, for all animals ($n = 7$) for stimuli presented at 60 dB (black) and 40 dB (blue) SPL. b) Average d' in quiet for same analysis. Error bars/shading represent the standard error of mean across animals.

confusion. There was some confusion within CvC sets for 20 dB SPL stimuli, suggesting the vowel component was more poorly encoded than the consonant component of the CvCs.

The addition of noise increased confusion in all conditions. As in quiet, the majority of confusion for 20 dB SPL stimuli was within CvC sets, suggesting again there was more difficulty encoding the vowel than the consonant components. The pattern of confusion for 40 dB SPL stimuli, while including some intra-set confusion, also demonstrated confusion between CvC sets. In particular, there was marked confusion between HvPs and HvDs. This trend towards increased confusion between CvC sets, rather than within sets, continued with stimuli at 60 dB SPL.

4. Discussion

We have presented analyses of data from multiunit activity recorded across the IC from ($n = 7$) normal-hearing cats that demonstrate that the IC encodes sufficient information to allow discrimination of speech stimuli over a range of stimulus intensities in quiet and moderate to loud stimuli in background noise. Furthermore, we have shown that neural

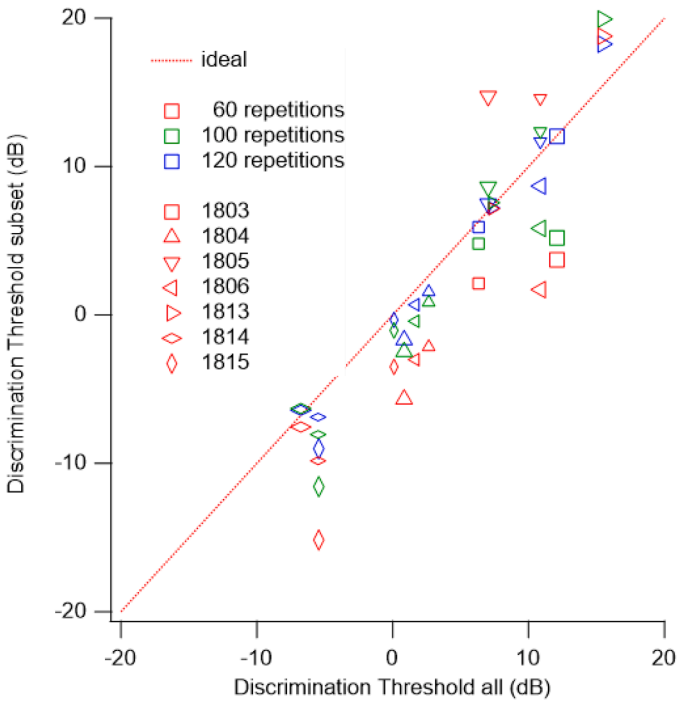

Fig. 7. Effective number of repetitions for individual cats ($n = 7$), as indicated (10ms bin width; BvDs only). Responses at 40 and 60 dB are indicated by decreasing size of symbols in the plot. The x axis shows the Discrimination Threshold for all stimuli presentations (~150) and the y axis shows the Discrimination Threshold for a subset of the stimuli presentations.

discrimination thresholds in noise are best when 1–2 ms bin widths are used for the analysis, but neural discrimination in quiet is maximized with an analysis bin width of 5–10ms.

The most comparable study in the literature is that of Perez et al. (2013) who recorded single and multiunit activity from a number of sites in the IC of each of 27 anesthetized rats and then combined responses across all animals prior to performing their neural discrimination analysis. Their stimuli were all presented in quiet, and they found that spike timing was particularly important for consonant encoding. Key differences between the Perez et al. (2013) study and the current study include that their stimuli were frequency shifted by one octave to better match the rat hearing range, while our stimuli were not modified (given the similar hearing range of cat and human in the speech range). Furthermore, Perez et al. combined recordings across animals, while we utilised recording from individual animals. Both of these differences are likely to have resulted in better preservation of temporal cues in our recordings (no potential for loss of temporal cues with frequency shifting and no blurring of timing across animals). However, our use of multi-unit recordings and a rate-based neural discrimination metric make definitive statements about the roles of rate versus temporal coding difficult as a well-vetted metric for the temporal analysis that is independent of rate is required for strong claims about rate vs timing coding. (Satuvuori and Kreuz, 2018). However, it is unclear to us that such a measure that incorporates multi-unit, multi-channel data (e.g. data containing both within and across channel temporal cues) is readily available. As our metric is a variant of the van Rossum distance, but computed over all recording channels simultaneously, we have confirmed that the average rate from CvC to CvC varied by approximately 30 %, which is within the range that should only be moderately affected by rate changes. We chose to focus on multi-unit recordings to maximise the probability of usable data on every electrode in the recording array and minimise any frequency sampling biases.

A novel aspect of the current study is the comparison of discrimination in quiet and in noisy conditions. Both measures of discrimination show a similar trend of poorer discrimination with decreasing stimulus intensity and larger analysis bin widths (above 20 ms). Where the two measures differ is the effects of using very short bin widths (<2 ms).

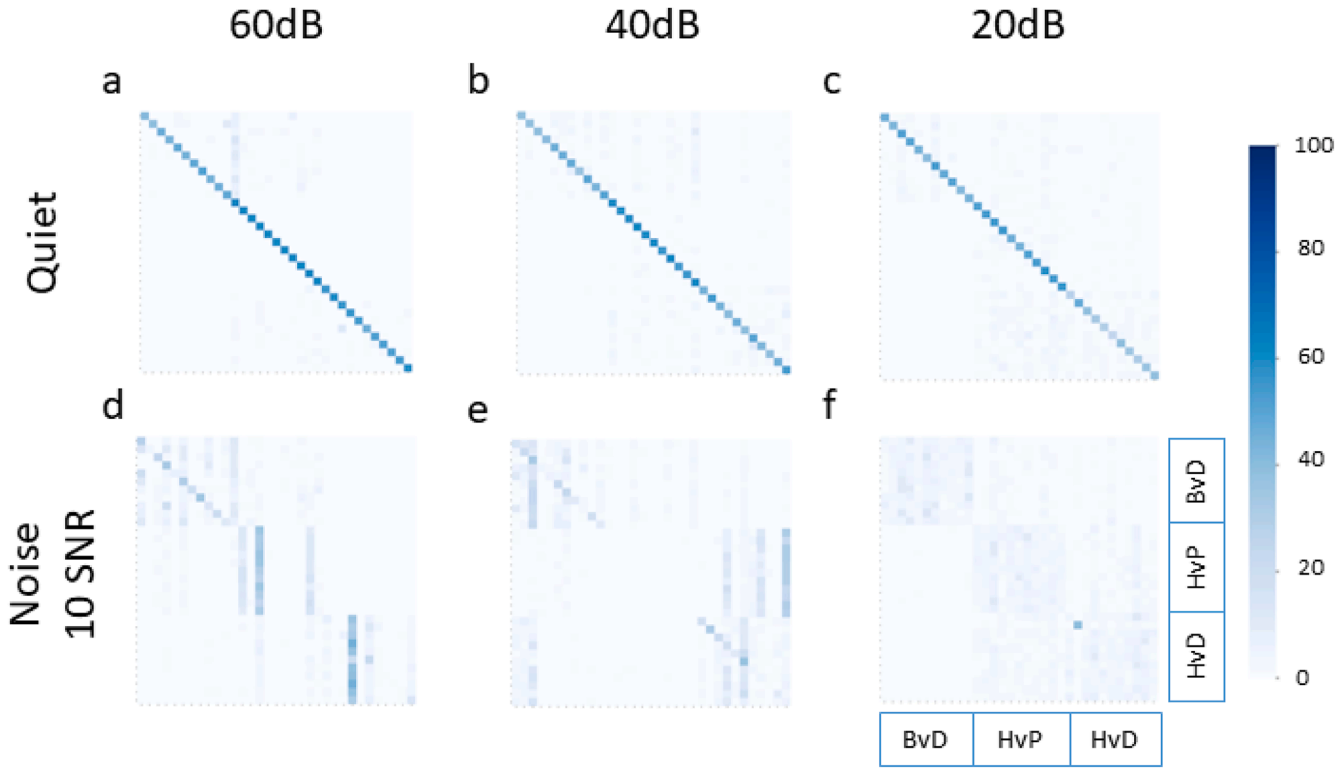

Fig. 8. A representative image of Confusion Matrices plotting all the CvCs for $n = 1$ animal (#1806) processed at a bin width of 10 ms in quiet and in 10 dB SNR for stimuli presented at 60 dB, 40 dB and 20 dB SPL.

Discrimination in quiet shows a distinct maximal discrimination around 5–10 ms, while discrimination threshold in noise continues to improve with bin width <2 ms, although performance at 0.1 ms is compromised, likely by excessive numbers of 'empty bins'. The difference between the two may be at least partially explained by the shorter bin widths allowing for a greater number of bins in which the signal dominates the noise. Our data likely underestimates this effect as we utilised broadband noise resulting in a relatively consistent level of noise, while the utilisation of babble noise, with its relatively more temporally variable profile, would likely lead to more analysis bins in which the signal dominates the noise.

One of the limitations of the methods utilised in this study is the time required to collect the neural responses for a set of CvC tokens. The recordings for a single set of CvC tokens consisting of 100+ repeats typically took 24 h or longer. However, for stimuli presented at 40 or 60 dB SPL, a minimum of 60 repeats was required. Therefore, if shorter protocols (60 repeats) had been used, recording durations would have been reduced by up to 50 %, allowing the testing of either additional speech tokens, or repeating with different modes of stimulation (i.e., as will be required to compare the effects of electric-only and combined electric and acoustic stimulation with a cochlear implant).

Another limitation of this study was that the neural responses were collected in response to speech tokens delivered by a single female speaker presented in quiet or broadband noise. We chose these conditions to emphasise the effects of energetic masking as these are most relevant for our anaesthetised cat model. The fundamental frequency of the female speaker is around the optimal processing window of 5–10 ms found in this study. Therefore, the optimal processing window might be dependent on the speaker's fundamental frequency. Future studies should look at input stimuli/CvCs spoken by different speakers so that the optimal processing window can be confirmed to be independent of speakers with different fundamental frequencies. Including both female and male speakers and various type of background noise would be of interest for future studies, particularly as previous research on neural recordings from the auditory cortex have shown that they were predictive of the pitch of the presented stimuli (Bizley et al., 2013; Gander et al., 2019). Furthermore, a recent study of recordings from the superior temporal gyrus showed that the auditory neuronal response to speech sounds may depend on external cues like acoustic-phonetic features rather than just the consonants and vowels involved (Leonard et al., 2024).

The confusion matrix analysis, particularly for louder sounds, suggested there were particularly speech tokens that were more readily confused than others (vertical stripes). This pattern is not observed in human speech discrimination and is likely a consequence of the analysis used, rather than a true representation of neural confusion. In particular, the confusions were driven by specific tokens that were represented furthest from the origin of our n-dimensional space (e.g., closest to a neurogram containing maximum spikes counts in all bins). The effects of different distance measures and other classification methods on the confusion matrix would be of interest for future studies.

Despite the limitations noted above, we can conclude that the described technique produces results that are similar to human psychophysics for similar speech discrimination tasks. This methodology therefore provides the basis for us to address the physiological basis for the benefits provided by EAS, and the interactions between the activity evoked by acoustic and electrical stimulation in the central auditory system. For example, the different optimal bin-widths for discrimination in quiet (5–10 ms) and noise (1–2 ms) may have a direct impact on clinical performance as cochlear implant and hearing aids are may present sound with different delays, effective smearing temporal cues over times, which these results would predict would be worse in noise than in quiet.

5. Conclusion

The present results indicate that, in normal hearing cats, the ability to discriminate speech tokens depends on the stimulus intensity and the levels of background noise. There is an optimal processing window between 5 and 10 ms for speech discrimination at all stimulus intensities.

Declaration of competing interest

None declared.

CRediT authorship contribution statement

Anu Sabu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Dexter Irvine:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization. **David B. Grayden:** Writing – review & editing, Funding acquisition, Conceptualization. **James Fallon:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Acknowledgements

This research was supported by the Australian Research Council's Discovery Project Grant (DP210103791). We would like to acknowledge the Passe and Williams Foundation for the Junior Fellowship Grant. The Bionics Institute also acknowledges the support they receive from the Victorian Government through its Operational Infrastructural Support Program. We would like to thank the Engineering Team and the Animal Research Team at Bionics Institute for their help with electrophysiological recordings and animal maintenance.

Data availability

Data will be made available on request.

References

Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., Schnupp, J.W.H., 2013. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr. Biol. 23 (7), 620–625. https://doi.org/10.1016/j.cub.2013.03.003.

Cantore, I., Santandrea, A., Carelli, G., De Canio, C., Oliveto, G., Piedilato, M.E., Valente, P., Santandrea, L., 2020. Electroacoustic cochlear stimulation advantages in severe-profound hearing loss. Hearing Balance Commun. 18 (2), 122–126. https://doi.org/10.1080/21695717.2019.1644860.

Chandrasekaran, B., Kraus, N., Wong, P.C.M., 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107, 1325–1336. https://doi.org/10.1152/jn.00923.2011.-A.

Fallon, J.B., Irvine, D.R.F., Shepherd, R.K., 2009. Cochlear implant use following neonatal deafness influences the cochleotopic organization of the primary auditory cortex in cats. J. Comparat. Neurol. 512 (1), 101–114. https://doi.org/10.1002/cne.21886.

Gander, P.E., Kumar, S., Sedley, W., Nourski, K.V., Oya, H., Kovach, C.K., Kawasaki, H., Kikuchi, Y., Patterson, R.D., Howard, M.A., Griffiths, T.D., 2019. Direct electrophysiological mapping of human pitch-related processing in auditory cortex. NeuroImage 202. https://doi.org/10.1016/j.neuroimage.2019.116076.

Gantz, B.J., Dunn, C.C., Oleson, J., Hansen, M.R., 2018. Acoustic plus electric speech processing: long-term results. Laryngoscope 128 (2), 473–481. https://doi.org/10.1002/lary.26669.

George, S.S., Wise, A.K., Fallon, J.B., Shepherd, R.K., 2015. Evaluation of focused multipolar stimulation for cochlear implants in long-term deafened cats. J. Neural Eng. 12 (3). https://doi.org/10.1088/1741-2560/12/3/036003.

Gifford, R.H., Davis, T.J., Sunderhaus, L.W., Menapace, C., Buck, B., Crosson, J., Neill, L. O., Beiter, A., & Segel, P. (2018). *Speech Understanding and Perceived Listening Difficulty* (Vol. 38, Issue 5). https://doi.org/10.1097/AUD.0000000000000418.Combined.

Heffer, L.F., Fallon, J.B., 2008. A novel stimulus artifact removal technique for high-rate electrical stimulation. J. Neurosci. Methods 170 (2), 277–284. https://doi.org/10.1016/j.jneumeth.2008.01.023.

Incerti, P.V., Ching, T.Y.C., Cowan, R., 2013. A systematic review of electric-acoustic stimulation: device fitting ranges, outcomes, and clinical fitting practices. Trends Amplif. 17 (1), 3–26. https://doi.org/10.1177/1084713813480857.

Irving, S., Wise, A.K., Millard, R.E., Shepherd, R.K., Fallon, J.B., 2014. A partial hearing animal model for chronic electro-acoustic stimulation. J. Neural Eng. 11 (4). https://doi.org/10.1088/1741-2560/11/4/046008.

Johnson, K.L., Nicol, T., Zecker, S.G., Bradlow, A.R., Skoe, E., Kraus, N., 2008. Brainstem encoding of voiced consonant-vowel stop syllables. Clin. Neurophysiol. 119 (11), 2623–2635. https://doi.org/10.1016/j.clinph.2008.07.277.

Leonard, M.K., Gwilliams, L., Sellers, K.K., Chung, J.E., Xu, D., Mischler, G., Mesgarani, N., Welkenhuysen, M., Dutta, B., Chang, E.F., 2024. Large-scale single-neuron speech sound encoding across the depth of human cortex. Nature 626 (7999), 593–602. https://doi.org/10.1038/s41586-023-06839-2.

Noh, H., Lee, D.H., 2012. Binaural electric-acoustic interactions recorded from the inferior colliculus of guinea pigs: the effect of masking observed in the central nucleus of the inferior colliculus. Clin. Exp. Otorhinolaryngol. 5 (3), 122–131. https://doi.org/10.3342/ceo.2012.5.3.122.

Perez, C.A., Engineer, C.T., Jakkamsetti, V., Carraway, R.S., Perry, M.S., Kilgard, M.P., 2013. Different timescales for the neural coding of consonant and vowel sounds. Cerebral Cortex 23 (3), 670–683. https://doi.org/10.1093/cercor/bhs045.

Portfors, C.V., Sinex, D.G., 2005. Coding of communication sounds in the inferior colliculus. In: Winer, J.A., Schreiner, C.E. (Eds.), The Inferior Colliculus. Springer, New York, pp. 411–425.

Presacco, A., Simon, J.Z., Anderson, S., 2019. Speech-in-noise representation in the aging midbrain and cortex: effects of hearing loss. PLoS ONE 14 (3). https://doi.org/10.1371/journal.pone.0213899.

Sato, M., Baumhoff, P., Tillein, J., Kral, A., 2017. Physiological Mechanisms in Combined Electric-Acoustic Stimulation. Otol. Neurotol. 38 (8), e215–e223. https://doi.org/10.1097/MAO.0000000000001428.

Satuvuori, E., Kreuz, T., 2018. Which spike train distance is most suitable for distinguishing rate and temporal coding? J. Neurosci. Methods 299, 22–33. https://doi.org/10.1016/J.JNEUMETH.2018.02.009.

Sinex, D.G., & Chen, G.-D. (2000). *Neural responses to the onset of voicing are unrelated to other measures of temporal resolution*. http://acousticalsociety.org/content/terms.

Vollmer, M., Hartmann, R., & Tillein, J. (2010). *Neuronal Responses in Cat Inferior Colliculus to Combined Acoustic and Electric Stimulation* (pp. 61–69). https://doi.org/10.1159/000262597.

Warrier, C.M., Abrams, D.A., Nicol, T.G., Kraus, N., 2011. Inferior colliculus contributions to phase encoding of stop consonants in an animal model. Hear. Res. 282 (1–2), 108–118. https://doi.org/10.1016/j.heares.2011.09.001.

Wilson, B.S., 2010. Partial deafness cochlear implantation (PDCI) and electric-acoustic stimulation (EAS). Cochlear Implants Int 11 (Suppl 1), 56–66. https://doi.org/10.1179/146701010x12671178390870.