Brochier, T., C. McKay, and H. McDermott. 2018. Encoding speech in cochlear implants using simultaneous amplitude and rate modulation. *The Journal of the Acoustical Society of America*. **144**(4): 2042-2051.

doi: 10.1121/1.5055989

Encoding speech in cochlear implants using simultaneous amplitude and rate modulation

Running Head:

Cochlear implants: amplitude and rate modulation

Tim Brochier,[1, a] Colette McKay,[2, b] Hugh McDermott[2, b]

[1]*University of Melbourne, Department of Medical Bionics, 384-388 Albert Street, East Melbourne, Victoria 3002, Australia*

[2]*The Bionics Institute, 384-388 Albert Street, East Melbourne, Victoria 3002, Australia*

[a] *Also at The Bionics Institute, 384-388 Albert Street, East Melbourne, Victoria 3002, Australia*

[b] *Also at University of Melbourne, Department of Medical Bionics, 384-388 Albert Street, East Melbourne, Victoria 3002, Australia*

To improve speech perception for cochlear implant (CI) users, it is essential to improve the transmission of temporal envelopes. The most common speech processors deliver temporal envelopes via the CI using fixed-rate amplitude modulated (AM) pulse trains. Psychophysical studies suggest that rate modulation (RM) and AM are perceived by a shared temporal integration mechanism, but the potential for them to constructively combine to encode temporal envelopes has yet to be explored. In this experiment, a speech processing strategy called ARTmod (Amplitude and Rate Temporal modulation) was developed to encode speech temporal envelopes with simultaneous AM and RM. The strategy was tested for perception of clean speech at 60 dBA and 40 dBA, and 60 dBA speech in noise (+10 dB SNR). The amount of RM was varied and the amount of AM was held constant to determine whether the addition of RM could enhance the perception of temporal envelopes and improve speech understanding. At the lowest RM amount, speech scores were poorest for all speech conditions. For 60 dBA clean speech and speech in noise, speech scores were significantly better at the highest RM amounts, suggesting that RM combined with AM can be used to enhance perception of temporal envelopes.

I. INTRODUCTION

Cochlear implants are a treatment option for people with severe to profound sensorineural hearing loss. Due to the low spectral resolution of the implant, CI users are especially reliant on temporal information to comprehend speech. Typically, CI speech processors encode temporal information by activating each implanted electrode with a fixed rate biphasic pulse train whose current level is modulated by the temporal envelope of the assigned acoustic frequency band. In this study, we evaluated a speech processing strategy that utilized simultaneous amplitude modulation (AM) and rate modulation (RM) to encode temporal envelopes of speech in quiet and in noise. The amount of RM was varied to determine whether the addition of RM can enhance the temporal speech cues and improve speech perception.

Low frequency temporal envelope cues from a limited number of spectral channels are sufficient for speech understanding (Azadpour and McKay, 2014; Fishman *et al.*, 1997; Friesen *et al.*, 2001; Nie *et al.*, 2006), with the most important temporal cues occurring at frequencies below 20 Hz (Arai and Greenberg, 1998; Fu and Shannon, 2000; Fu and Galvin III, 2001; Saberi and Perrott, 1999; Shannon *et al.*, 1995; Shannon, 2002; Vandali *et al.*, 2000). Several psychophysical studies have shown that the ability of a CI user to detect low frequency amplitude modulations is correlated with their speech perception ability (Brochier *et al.*, 2017; De Ruiter *et al.*, 2015; Fu, 2002; Garadat *et al.*, 2012; 2013; Gnansia *et al.*, 2014; Luo and Fu, 2007; Won *et al.*, 2011). Limited access to low frequency temporal amplitude modulation cues may be one reason for poor speech outcomes for some CI users.

Speech processing strategies that enhance low frequency amplitude modulations have led to improved speech perception for CI users (Azadpour and Smith, 2016; Geurts and Wouters,

1999; Holden *et al.*, 2005; Koning and Wouters, 2016; Vandali, 2001). In general, these strategies detect onsets or peaks in the temporal envelope of each acoustic frequency band, and amplify those segments by increasing the current level.

An electrical hearing model has been developed to explain the temporal processing capabilities of CI users (McKay and McDermott, 1998; McKay *et al.*, 2001; McKay *et al.*, 2003; McKay *et al.*, 2013). A key component of this model is central temporal integration, where the neural activity elicited by each pulse is summed over time by a sliding temporal integration window with an equivalent rectangular duration of 7 ms. The weighting of the window is determined by asymmetrical exponential functions that account for forward and backward masking, similar to acoustic hearing models (Oxenham and Moore, 1994; Oxenham, 2001; Plack *et al.*, 2002). Assuming that the output of the temporal integration window is used to encode envelope information, and because both RM and AM result in modulations at the output of the temporal integration window, both RM and AM could be used to enhance perception of low frequency temporal envelope information.

Experiments investigating RM and AM on single electrodes support the idea that both modulation types may be detected by a common temporal integration mechanism, and that RM could be used to elicit similar percepts to AM. Luo and Fu (2007) measured rate modulation detection thresholds (RMDTs) at a modulation frequency of 10 Hz with and without simultaneous AM interference at a different modulation frequency. RMDTs were significantly worse with simultaneous noisy AM, and with simultaneous sinusoidal AM at a modulation frequency of 20 Hz. These results showed that RM and AM interact with each other, suggesting that a common temporal integration mechanism may underlie RM and AM detection. Brochier *et al.* (2018) measured RMDTs and AMDTs in CI users. RMDTs and AMDTs were correlated, and

both were best at low modulation frequencies and high presentation levels, providing further evidence that RM and AM may be perceived by a common temporal integration mechanism.

The signal processing strategy presented in this paper, ARTmod (Amplitude and Rate Temporal modulation), encoded low frequency temporal envelopes with simultaneous AM and RM. The strategy encoded an additional temporal cue in the pulse trains on each electrode, with both AM and RM encoding the acoustic low-frequency envelope in each channel. This implementation of RM was not intended to encode temporal pitch information, as has been investigated in other studies (Arnoldner *et al.*, 2007; Magnusson, 2011; Nie *et al.*, 2005). Instead, ARTmod used RM and AM to encode the magnitude of the low frequency temporal envelope in each channel, where most of the useful speech information is encoded in temporal modulations below 20 Hz. It was hypothesized that the additional RM encoding would enhance the perception of these low frequency envelopes in the speech signal, and hence improve speech perception.

The purpose of this experiment was to extend the investigation of RM perception from single-electrode sinusoidal envelopes (Brochier *et al.*, 2018) to multiple-electrode speech envelopes. The effect of RM was evaluated by measuring speech perception with an implementation of the ARTmod processing strategy that encoded speech envelopes using various amounts of RM in addition to a fixed amount of AM. Speech scores were measured in quiet at presentation levels of 60 dBA and 40 dBA, and in multi-talker babble noise at 60 dBA. It was hypothesized that speech scores would improve as the amount of RM was increased.

II. METHODS

A. Participants

Eight postlingually deafened adult CI users completed the study. Participants were recruited from the clinical population of the Royal Victorian Eye and Ear Hospital in Melbourne. Approval to conduct the studies was obtained from the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital, and each participant provided written informed consent. Participants were tested over the course of three to four sessions of 2 hours. Details about the participants are described in *Table I.*

Table I. Relevant information about the CI users who participated in the study.

	Gender	Age (years)	Duration of hearing loss before implantation (years)	Duration of implant use (years)	Implant Type and Stimulation rate	Etiology
P1	Male	74	1	6	CI422 900 pps	Noise Exposure
P2	Male	75	42	7	CI24RE 500 pps	Unknown, progressive hearing loss
P3	Male	44	5	5	CI512 900 pps	Genetic
P4	Female	39	15	4	CI24RE 900 pps	Genetic
P5	Female	66	20	12	CI24M 900 pps	Cerebral Palsy
P6	Male	57	23	7	CI24RE 900 pps	Unknown, Genetic
P7	Female	65	16	5	CI24RE 900 pps	Unknown, progressive hearing loss
P8	Female	86	15	7	CI512 900 pps	Antibiotics for pneumonia

B. Equipment

All speech stimuli were pre-processed offline in MATLAB. The Nucleus MATLAB Toolbox, provided by Cochlear Corporation, was used to implement signal processing

procedures similar to the Cochlear Freedom CP900 series processor. The Nucleus Implant Communicator (NIC) streaming research platform, also provided by Cochlear Corporation, was used in conjunction with a Cochlear L34 Research Processor to deliver the processed stimuli to the participants.

C. Processing Strategy

Similar to most CI speech processing strategies, the ARTmod strategy converted a sampled audio signal to an electrical pulse sequence. A block diagram of the signal path is shown in Figure 1.

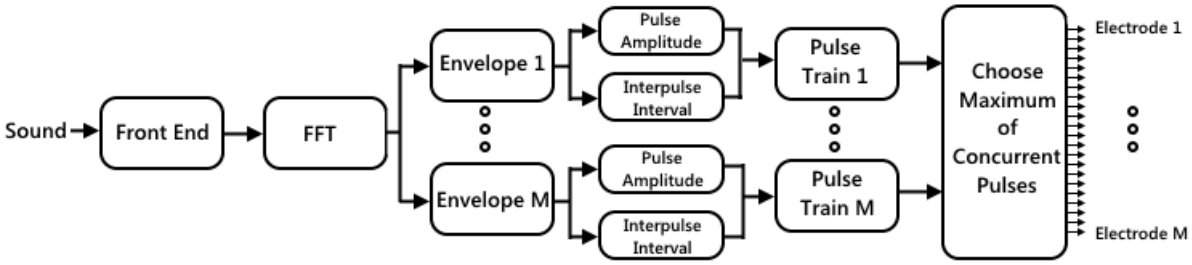


Figure 1. Block diagram for the ARTmod speech processing strategy.

The signal processing consisted of the following stages: front end processing, envelope extraction, pulse amplitude calculation, interpulse interval calculation, and pulse train collation. The front end processing, envelope extraction, and pulse amplitude calculation stages were the same as in the Advanced Combination Encoder (ACE) strategy (Wouters *et al.*, 2015). The unique feature of the ARTmod processing strategy was the interpulse interval calculation stage. Unlike fixed-rate speech processing strategies, where the stimulation rate is kept constant, the ARTmod strategy applied RM by varying the time interval following each pulse along with the

pulse amplitude according to the magnitude of the temporal envelope in each frequency band. The following sections describe each stage in detail.

Front End Processing

The front end processing consisted of a gain stage, a microphone filter, and an automatic gain control (AGC). The gain stage was calibrated to ensure that the ARTmod strategy was operating on the same envelope magnitudes as the ACE strategy at presentations levels of 60 dBA and 40 dBA. The calibrated gain value was determined using test tones at the center frequency of each electrode in the array, and verified using a recorded spoken sentence. The microphone filter function emulated the frequency response of the Cochlear Freedom microphone, and the AGC function was identical to the AGC of the CP910 processor. Both functions were provided by the Nucleus MATLAB Toolbox.

Envelope Extraction

The envelope extraction phase was the same as that used in the ACE scheme. For the incoming audio signal sampled at 30 kHz, the FFT was calculated at a rate of 15 kHz using a 128-point Hann window. The FFT magnitudes were summed in bins corresponding to center frequencies of 250, 375, 500, 625, 750, 875, 1000, 1125, 1250, 1438, 1688, 1938, 2188, 2500, 2875, 3313, 3813, 4375, 5000, 5688, 6500, 7438 Hz. Each of the 22 bins represented an electrode in the implanted array. Once the FFT magnitudes were summed in bins, an exponential function known as the loudness growth function (LGF) was applied to convert the calculated magnitude to a transformed envelope magnitude value between 0 and 1. The LGF was characterized by three clinical parameters: base level, saturation level, and Q. The values of these parameters were kept at the default values used in the ACE processing strategy, which were

4/256 for the base level, 150/256 for the saturation level, and 20 for the Q value. The base level and saturation level refer to the minimum and maximum acoustic levels in each frequency band that are encoded. Both are frequency-dependent and signal-dependent. Due to the frequency response of the front-end processing, the base level for sine tones was 13 dB SPL for channels 1 through 8 (center frequencies 7438 down to 2875 Hz), and rose to 30 dB SPL at a slope of approximately 6 dB/octave for channels 9 through 22 (center frequencies 2300 down to 250 Hz). The saturation level was 40 dB above the base level in each frequency band.

FFT bin magnitudes equal to the base level were assigned a transformed envelope magnitude of 0, and FFT bin magnitudes equal to or greater than the saturation level were assigned a transformed envelope magnitude of 1. The Q value determined the amount of curvature in the LGF, and FFT bin magnitudes between the base level and the saturation level were assigned a value between 0 and 1 accordingly. If the FFT bin magnitude at a particular sample was less than the base level, that sample was assigned a small negative value, as a flag to mark samples where no pulses should be assigned.

Interpulse Interval Calculation

RM was applied on each electrode separately by adjusting the interpulse interval following each pulse, which was calculated from the transformed envelope magnitude. In each frequency band, the first transformed envelope sample with a magnitude greater than or equal to 0 was assigned a pulse and an interpulse interval following that pulse. If the transformed envelope sample following that assigned interpulse interval was greater than or equal to 0, a pulse was assigned to that sample and another interpulse interval was calculated. If the transformed envelope sample following the assigned interpulse interval was less than 0 (marked

by the small negative value), a pulse was assigned to the next transformed envelope sample with a magnitude greater than or equal to 0, and another interpulse interval was calculated. This process was repeated through the duration of the stimulus.

For the calculation of interpulse intervals (the non-stimulating period between a pulse's offset and the following pulse's onset), a minimum and a maximum rate were chosen. The minimum rate (largest interpulse interval) was used for the lowest transformed envelope magnitude of 0, and the maximum rate (smallest interpulse interval) was used for the highest transformed envelope magnitude of 1. For envelope values between 0 and 1, the interpulse interval following each pulse was calculated according to Equations 1 and 2:

$$Rate = MinRate * \left(\frac{MaxRate}{MinRate}\right)^{Transformed\ Env\ Mag} \quad (1)$$

$$Interpulse\ Interval = \frac{1}{Rate} - BPD \quad (2)$$

BPD represents the biphasic pulse duration, the sum of the duration of the positive and negative phases of the biphasic pulse (both 20 μs) and the interphase gap (8.4 μs). The amount of RM was adjusted by changing the *MinRate* and *MaxRate* variables in Equation 1. The RM ratio was calculated as the ratio between *MaxRate* and *MinRate*. RM ratios of 1, 2, 4, and 10 were used. Table 2 shows the minimum and maximum rates used for the different RM ratios in this experiment. At an RM ratio of 1, the rates varied between 1176 pps and 1248 pps. This small amount of rate modulation was used for the ratio of 1 (rather than a fixed rate of 1200 pps) because there were too many concurrent pulses across channels when there was zero rate variation. At RM ratios of 2, 4, and 10 the rates varied between 900 and 1800 pps, 600 and 2400 pps, and 300 and 3000 pps, respectively. These RM ranges were chosen so that the pulse width

could be held constant at 20 µs across RM ratios, while not introducing a rate pitch for the low rates.

Table II. Minimum and maximum rates and interpulse intervals used in the different conditions for the ARTmod strategy.

RM Ratio	Min Rate (Max Interpulse Interval)	Max Rate (Min Interpulse Interval)
1	1176 pps (792 µs)	1248 pps (743 µs)
2	900 pps (1053 µs)	1800 pps (497 µs)
4	600 pps (1608 µs)	2400 pps (358 µs)
10	300 pps (3275 µs)	3000 pps (275 µs)

Pulse Amplitude Calculation

The electrical amplitude of each pulse, reported in clinical current-level units (CL steps), was calculated based on the transformed envelope magnitude. For the CI24M implant each CL step represents a 0.176 dB change in current, and for the CI24RE and CI512 implants each CL step represents a 0.157 dB change in current.

Amplitude values were mapped linearly in CL steps between the threshold level, measured with a fixed-amplitude pulse train at a rate of 500 pps ($T_{500\ pps}$), and the comfort level measured with a fixed-amplitude pulse train at a rate of 2400 pps ($C_{2400\ pps}$). $T_{500\ pps}$ and $C_{2400\ pps}$ levels were chosen to ensure that, when RM was included, the resulting stimuli were not too loud for intense stimuli, or inaudible for soft stimuli. Thus, the dynamic range in CLs was smaller than that present in a fixed-rate strategy at a stimulation rate of 1200 pps, resulting in an intentionally compressed AM representation of the temporal envelope (in current amplitude modulations) across RM conditions. The mapping equation, shown in Equation 3, is identical to that used in ACE, but with threshold and comfort levels derived from the two different rates:

$$Pulse\ Amplitude = T_{500\ pps} + \left(C_{2400\ pps} - T_{500\ pps}\right) \times Transformed\ Env\ Mag \qquad (3)$$

It is important to note that for every RM ratio used (every combination of *MinRate and MaxRate*), the same $T_{500\ pps}$ and $C_{2400\ pps}$ levels were used. In this way, the amount of RM was varied, while the amount of AM (current modulation) remained the same in each experimental condition. When there was no RM (RM ratio = 1), the centrally perceived envelope would be compressed relative to the case where T and C are both measured at 1200 pps. As the RM ratio was increased, it was hypothesized that the centrally-perceived modulation would be enhanced if AM and RM can constructively combine in the central neural pathways to encode temporal envelopes. In this way, it was proposed that the RM ratio of 4 (600 pps to 2400 pps) would centrally recover the perceived temporal envelope to be similar to a fixed-rate strategy, and the RM ratio of 10 (300 pps to 3000 pps) would enhance the envelope to a degree more than in a normal fixed-rate strategy. Figure 2 illustrates this concept with a 500 Hz sine wave that has been ramped up in level. The upper panel shows the original waveform, and the lower panel shows the output of the ARTmod strategy at electrode 20 (500 Hz center frequency) for two different RM ratios. Note that the amount of AM is equal while the amount of RM is varied for the different RM ratios.

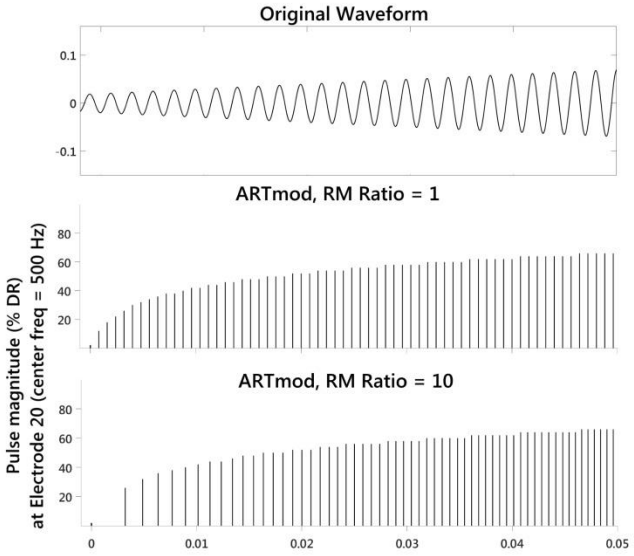


Figure 2. ARTmod outputs at different RM ratios for a 500 Hz ramped sinusoid. The upper panel shows the original waveform, and the lower panels show the ARTmod output at electrode 20 (center frequency 500 Hz). At the RM ratio of 1, there is no change in stimulation rate as the pulse amplitude increases, while at an RM ratio of 10, the stimulation rate rises as the pulse amplitude increases.

Pulse Train Collation

In the collation stage, the pulse trains in each frequency band were collated using the calculated pulse amplitudes and interpulse intervals. If there were any concurrent pulses (overlapping in time) on different electrodes, the pulse with the highest envelope magnitude value was chosen, and the lower magnitude temporally overlapping pulses were removed. Unlike ACE, no n-of-m maxima selection scheme was implemented before this pulse collation stage. N-of-m maxima selection was left out of the ARTmod processing to minimize additional modulations in the temporal envelopes based on whether a channel was one of the maxima or not. Figure 3 shows electrodograms displaying the sequence of pulse amplitudes over time at

each active electrode for the ACE processing strategy and the ARTmod strategy at RM ratios of 1, 2, 4 and 10.

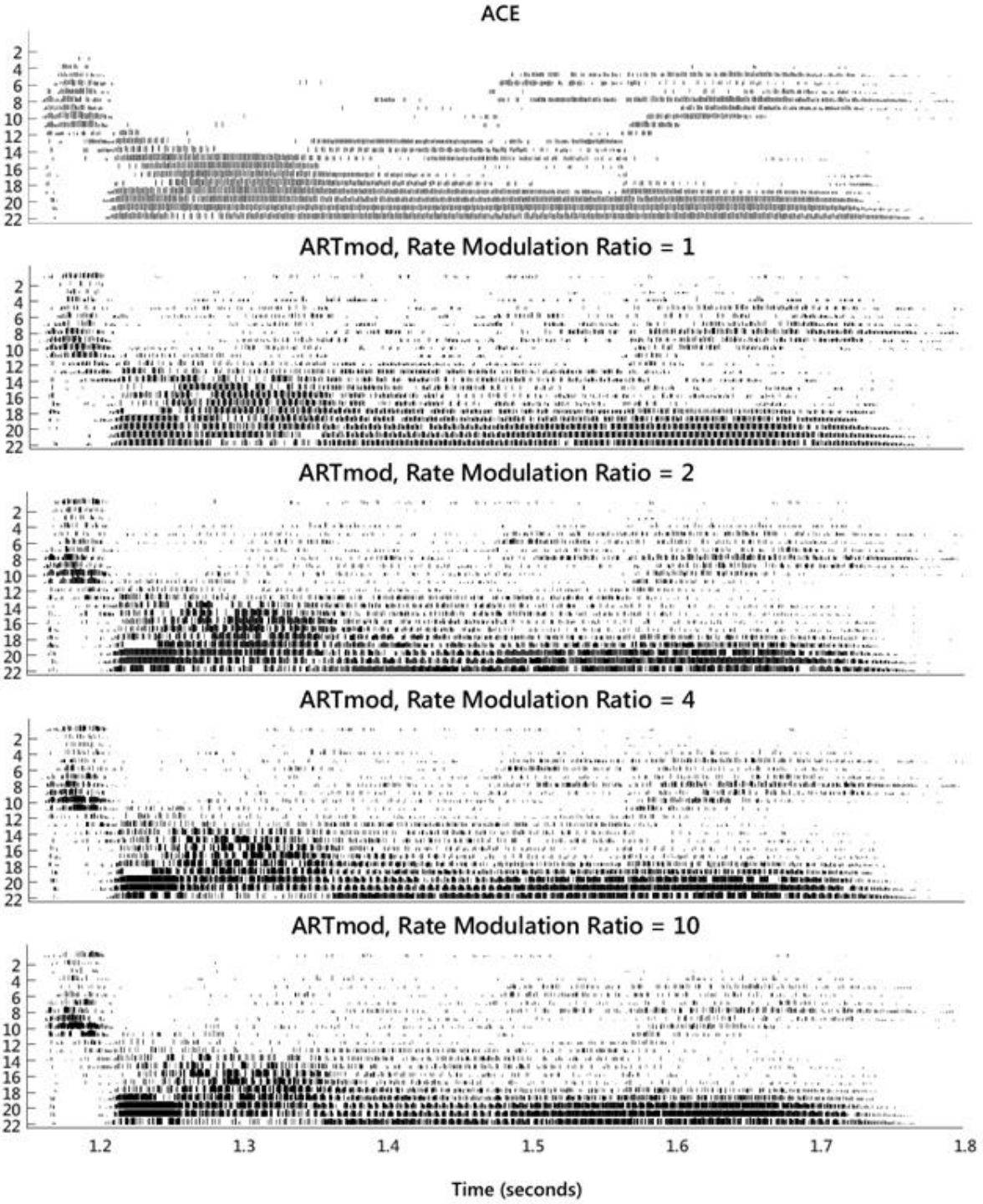


Figure 3. Electrodograms displaying the sequence of pulse amplitudes over time at each active electrode. Each stimulation sequence represents the word 'STRONGLY' spoken at a presentation level of 60 dBA. The first panel was encoded using the ACE strategy, and panels 2 through 5 were encoded using the ARTmod strategy with an RM ratio of 1, 2, 4, and 10, respectively.

Electrodograms of each RM ratio for 40 dBA clean speech and 60 dBA speech in babble noise at 10 dB SNR are provided in the supplementary materials.

D. Stimuli and Procedures

Mapping

The standard clinical adjustment procedure at the Royal Victorian Eye and Ear Hospital was used to find T-levels for the 500 pps rate and C-levels for the 2400 pps rate. The balance across electrode positions was checked by comparing loudness in groups of three electrodes, using the Custom Sound fitting software.

Conditions

The ARTmod strategy was tested in 12 different conditions. Speech scores were obtained for 60 dBA and 40 dBA clean speech, and 60 dBA speech in +10 dB SNR multi-talker babble noise. In each of the three speech conditions, four RM ratios were used. The RM ratios tested were 1, 2, 4, and 10. The order of the testing of the four RM ratios was varied using a 4 by 4 Latin Square design. In addition, to provide information for the reader about each participant's clinical performance, their speech scores were obtained using the same speech material and their existing clinical MAP.

Loudness Balancing

Stimuli with RM ratios of 1, 2, and 10 were all loudness-balanced to speech stimuli with an RM ratio of 4, so that each of the stimuli for the ARTmod strategy was approximately the same overall loudness in each RM condition. The stimuli used to balance loudness were BKB sentences. A method of adjustment was used in which participants were asked to raise and lower the level of a test stimulus until it matched the loudness of the reference stimulus. The loudness balancing software raised and lowered the level by adding or subtracting equal CL steps to all pulses in the pre-processed test stimulus. Two rounds of loudness balancing were performed for each combination at 60 dBA and 40 dBA. In the first round, the reference MAP used an RM ratio of 4, and the other conditions were adjusted in level. In the second round, the other conditions were used as the reference MAPs and the stimuli with an RM ratio of 4 were adjusted in level. The loudness balance adjustment was determined by averaging the balanced levels from the two rounds. It was found that the average loudness balance adjustment was less than one CL step for every RM ratio, indicating that the overall loudness difference between RM conditions was minimal.

The RM ratio of 4 was loudness balanced to the reference ACE MAP at 60 dBA and 40 dBA using the same procedure. Every participant needed their ARTmod stimuli reduced in level compared to ACE. At 60 dBA, the mean loudness balance adjustment was -6.250 ± 3.012 CL steps. At 40 dBA, the mean loudness balance adjustment was -8.125 ± 2.588 CL steps. The adjustments were most likely necessary due to the high overall maximum implant stimulation rate of 15000 pps, compared to that of ACE with 8 maxima, which is 7200 pps (at 900 pps per channel) or 4000 pps (at 500 pps per channel). Because there were only small loudness

differences measured between RM ratios (less than one CL on average), these adjustments were applied to all ARTmod stimuli at all RM ratios.

Speech Perception

Speech perception was evaluated using BKB sentences, recorded by an Australian female speaker. They comprised 30 lists of 16 sentences, each sentence containing 3 or 4 keywords, for a total of 50 keywords per list. All stimuli were processed offline, converted to electrical pulses, and streamed to the implant using the NIC. Each of the 15 conditions was tested with two lists, and the scores were recorded as number of keywords correct out of 100.

III. RESULTS

A. Effect of Rate Modulation Amount

Sentence scores for each subject with each RM ratio are shown for 60 dBA clean speech, 40 dBA clean speech, and 60 dBA speech in speech babble noise (Figure 4). The speech score data were negatively skewed, with many subjects scoring above 90% correct. Because the data were not normally distributed, an arcsine square root transformation was applied to all of the data before a statistical analysis was carried out.

Separate one-way repeated measures ANOVAs were used to assess the effects of RM amount for each speech condition. The ANOVA revealed a significant effect of RM amount for 60 dBA clean speech ($F(1,7) = 4.24$, $p = 0.017$) and 60 dBA speech in multi-talker babble noise ($F(1,7) = 9.79$, $p < 0.001$). A Tukey pairwise comparison was used to assess the specific effects of different RM ratios, finding that speech scores generally increased as the RM ratio increased in these speech conditions. For 60 dBA clean speech, speech scores with an RM ratio of 10 were

311 significantly better than speech scores with an RM ratio of 1 ($p = 0.009$). Scores for 60 dBA
312 speech in multi-talker babble noise with RM ratios of 10 and 4 were significantly better than
313 scores with an RM ratio of 1 ($p = 0.001$ for both). Scores with an RM ratio of 4 were not
314 significantly better than with an RM ratio of 2 ($p = 0.053$). For 60 dBA clean speech, participants
315 with the lowest average scores (S1, S4, and S5) showed the most benefit of increased RM
316 amount, and each of them had their highest score at an RM ratio of 10. For the other subjects,
317 speech scores were very high for 60 dBA clean speech (>95% across RM ratios), so differences
318 related to RM ratios were minimal. All of the participants obtained their highest ARTmod
319 strategy scores at RM ratios of either 4 or 10 for 60 dBA speech in noise. At 40 dBA the effect of
320 RM amount was not significant ($F(1,7) = 1.29$, $p = 0.305$).

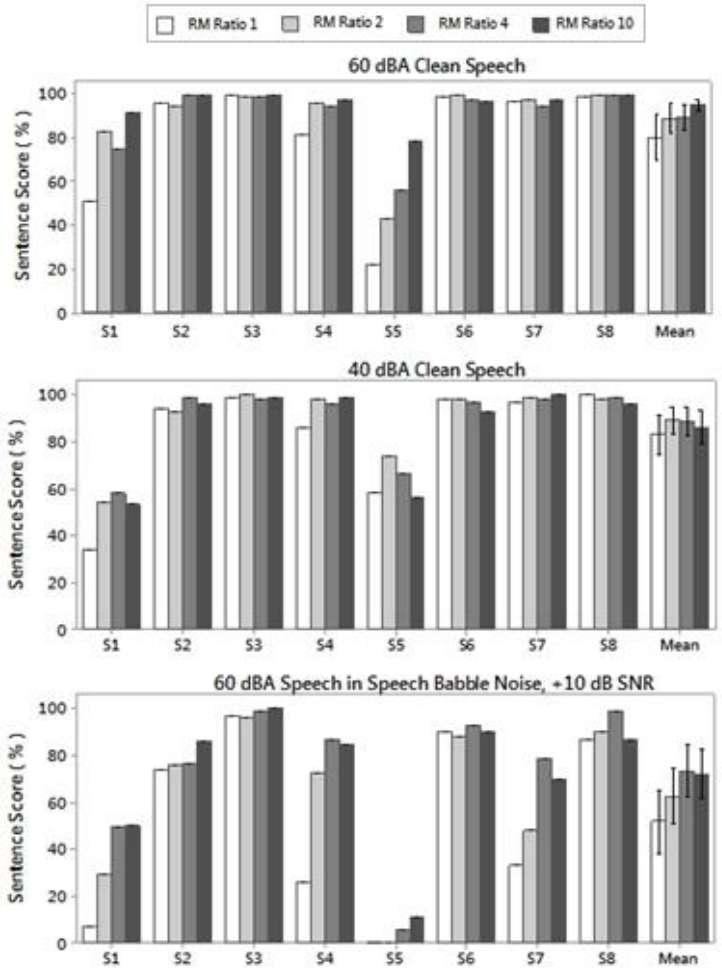


Figure 4. Speech scores for 60 dBA clean speech (upper panel), 40 dBA clean speech (middle panel), and 60 dBA speech in speech babble noise (lower panel) for each participant in each condition. For the mean scores, the error bars represent one standard error of the mean.

Figure 5 shows how the different ARTmod conditions, which were tested acutely, compared with the ACE processing scheme, which was used by the participants on an everyday basis. ARTmod enabled speech perception in the same range as ACE for speech in quiet conditions, but not for speech in noise.

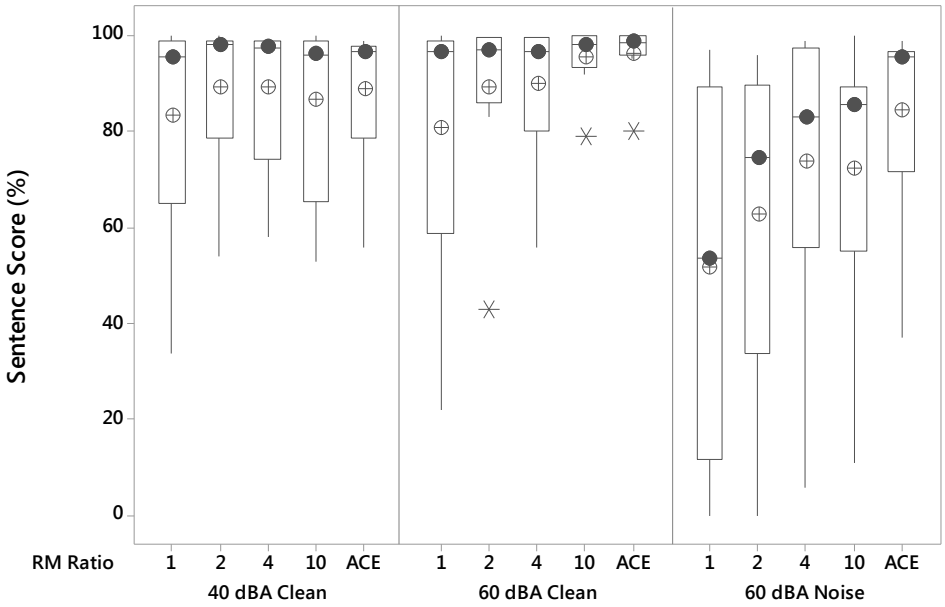


Figure 5. Boxplot of speech scores across participants for the ARTmod strategy with RM ratios of 1, 2, 4, and 10, and their normal ACE MAP. The left panel represents the scores for 40 dBA clean speech, the middle panel represents the scores for 60 dBA clean speech, and the right panel represents scores for 60 dBA speech in speech babble noise (+10 dB SNR). Boxes represent the interquartile ranges, crossed open circles represent the means, filled black circles represent the medians, and asterisks represent outliers.

IV. DISCUSSION

A. Effect of Rate Modulation Amount

A speech processing strategy that used AM and RM to encode temporal speech envelopes, ARTmod, was devised and assessed. As hypothesized, a significant effect of RM amount was found for 60 dBA speech in quiet and in speech-babble noise, with speech scores improving as the RM ratio increased. For 60 dBA speech in noise, average speech scores across

subjects were significantly better at RM ratios of 4 and 10 than they were at the RM ratio of 1. For 60 dBA speech in quiet, average speech scores across subjects were significantly better at the RM ratio of 10 than they were at the RM ratio of 1. For all conditions, speech scores were lowest at the RM ratio of 1. The improvement with increased RM ratios indicates that RM was effective in enhancing the participants' perception of temporal envelope cues.

The perceptual mechanism underlying the beneficial effect of RM most likely involved central temporal integration, whereby RM and AM could combine to expand the temporal envelopes at a perceptual level. Luo and Fu (2007) demonstrated that AM interfered with RM detection when different modulation frequencies were used in the two modulations, inferring that both were perceived via the same central mechanism. Brochier *et al.* (2018) showed that AM and RM were correlated and were affected similarly by modulation frequency and presentation level, supporting the idea of a shared mechanism for AM and RM detection. The present study also supports this conclusion by showing that AM and RM can be combined to improve the perception of temporal envelopes in speech. For speech at 40 dBA, no significant effect of RM amount was found. The degradation of scores at RM ratio 10 for the low level of 40 dBA may be due to the use of 300 pps as the minimum rate, below the 500 pps rate at which the T_{500pps} level was measured. Consequently, some low-level envelope cues may have been inaudible for some participants.

While this study demonstrated that AM and RM cues can constructively combine, it did not empirically prove that CI users were able to make use of the additional interpulse interval cues provided by RM. The implementation of ARTmod in the present study allowed the amount of RM to be varied while keeping the amount of AM fixed. This methodology necessitated compression of the AM representation of the temporal envelopes. It is plausible that expanding

366 the AM representation using additional AM (rather than RM) would lead to similar benefits as
367 observed in this experiment with AM + RM. Nevertheless, the results supported our hypothesis
368 that AM and RM constructively combine to encode temporal envelopes, which may inform
369 future CI speech processing strategies that utilize RM.

370 ## B. Effect of Concurrent Pulse Removal Stage

371 An analysis of the concurrent pulse removal stage was used to assess whether systematic
372 differences in pulse removal across RM ratios affected the results. In order to better characterize
373 the influence of the concurrent pulse removal stage on the results, the percentage of pulses
374 removed and effect of pulse removal on the spectral envelope were examined for each RM ratio.
375 Table III displays the percentage of pulses removed for all stimuli at each RM ratio after the
376 concurrent pulse removal stage.

377 **TABLE III. Percentage of pulses removed after concurrent pulse removal stage.**

	60 dBA Clean	40 dBA Clean	60 dBA + 10 dB SNR
RM 1	48.97%	34.15%	53.13%
RM 2	51.86%	33.05%	56.06%
RM 4	53.69%	30.76%	56.24%
RM 10	51.89%	24.70%	52.60%

378

379 For speech at 60 dBA in quiet and in noise (where the effect of RM amount was significant), the
380 percentages of pulses removed were similar across RM ratios, and there was not a monotonic
381 pattern relating RM ratio to the percentage of pulses removed. The analysis suggests that the
382 observed effects of RM amount were not influenced by the percentage of removed pulses during
383 the concurrent pulse removal stage. For the 40 dBA condition, the amount of removed pulses

decreased as the RM ratio increased. This result is intuitive, because at low levels and high RM ratios the overall rate dropped below the central rate of 1200 pps, so the probability of pulses overlapping was reduced. Despite having a smaller percentage of pulses removed, there was no effect of RM amount on speech perception in the 40 dBA condition.

In addition, the distribution of stimulation activity across the electrode array was evaluated for different RM ratios and speech stimuli. The distribution of stimulation activity was calculated from the transformed envelope magnitude (values between 0 and 1) of the pulses in the stimulation sequence (see section II C). Transformed envelope magnitude values were used, rather than the pulse amplitudes (in CL steps), to normalize for the various T_{500pps} and $C_{2400pps}$ levels of the individual subjects. The transformed envelope magnitude values of each pulse were summed across 16 sentence stimuli for each of the 22 individual channels. Those 22 values were divided by the sum across channels, resulting in a stimulation activity histogram that showed the average proportion of the total stimulation activity on each channel.

The stimulation activity histogram for 60 dBA clean speech is shown in Figure 6. Only this condition is shown because the pattern was similar across speech conditions. Stimulation activity histograms were similar for the different RM ratios. The primary difference was that as the RM ratio increased, more emphasis was placed on the most apical channels (particularly electrodes 19, 20, and 21). For example, with the RM ratio of 10, high-magnitude envelope samples on apical channels were represented by high pulse rates (up to 3000 pps), which increased the probability of pulses on those channels temporally overlapping with pulses on other channels. In the pulse train collation stage, pulses with relatively high magnitude were selected for delivery to the electrodes in preference to those representing lower magnitudes, resulting in an increased proportion of the total stimulation activity occurring on apical channels.

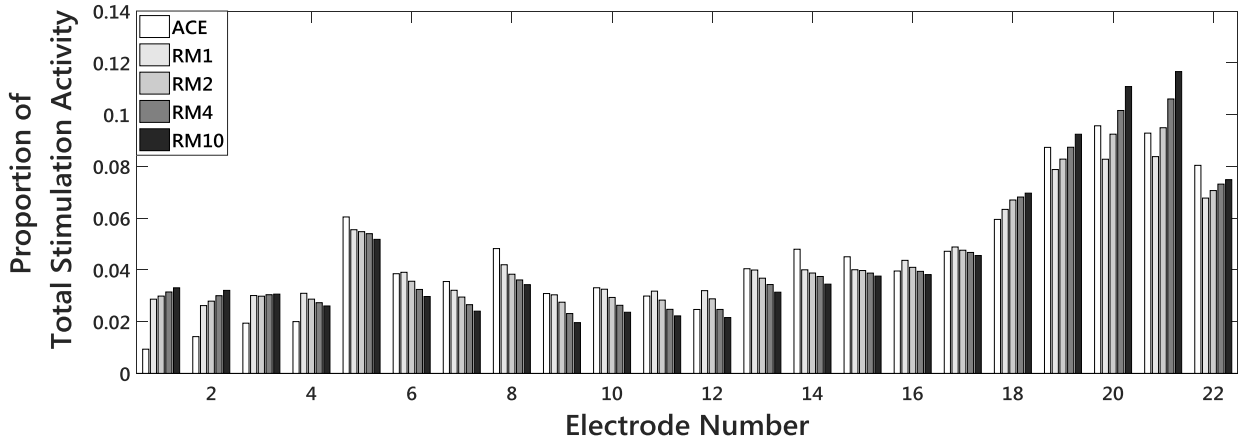


Figure 6. Stimulation activity histogram for 60 dBA speech, displaying the average proportion of stimulation represented by each electrode for the different RM ratios and ACE. Transformed envelope magnitude values were summed across 16 sentence stimuli for each individual electrode, and then those 22 values were divided by the sum across channels.

Further analysis was completed to assess whether this apical emphasis for the higher RM ratios was likely to have influenced the speech recognition scores. Weatherby *et al.* (2003) evaluated the effect of spectral pre-emphasis on speech understanding. They compared sentence scores in noise for high frequency (HF) emphasis, low frequency (LF) emphasis, and default sound-processing conditions. These emphasis filters were applied before the filterbank, so they resulted in a comparable effect to the present study, where the processing affected the selection of electrodes for stimulation. In order to compare their results with the present study, a stimulation activity histogram was created using the three pre-emphasis settings from Weatherby *et al.* (2003) for the same 16 sentences as in the above analysis (see supplementary materials). The differences in the distribution of stimulation activity between the default, HF pre-emphasis, and LF pre-emphasis applied by Weatherby et al (2003) were greater than they were in this study

between different RM ratios. Weatherby et al (2003) found that perception of sentences in noise was significantly worse with the LF pre-emphasis than it was with the HF pre-emphasis, and that neither the HF nor LF pre-emphasis conditions produced significantly different speech recognition scores than the clinical MAP with default pre-emphasis. Furthermore, the distribution of stimulation activity was similar between the LF pre-emphasis condition in the Weatherby et al (2003) study and the RM10 condition in the present study. Therefore, it is unlikely that the benefit observed with increased RM ratios in the present experiment was related to the relatively small increase in stimulation activity on apical channels that occurred with the higher RM ratios.

C. Comparison to other Envelope Enhancement Studies

The ARTmod strategy with an RM ratio of 10 theoretically expanded the temporal envelopes at the perceptual level compared to participants' normal MAPs. Other envelope enhancement strategies (Azadpour and Smith, 2016; Geurts and Wouters, 1999; Holden *et al.*, 2005; Koning and Wouters, 2016; Vandali, 2001) have shown benefit by increasing the current level at onsets or during transients of the envelope. The latter technique differs from the current implementation of the ARTmod strategy, which applied envelope enhancement to the entire envelope rather than a specific portion of it. A potential variation of the ARTmod strategy would be to derive the RM signal from portions of the amplitude envelope, or from other features of the amplitude envelope.

Another noteworthy difference between this study and other envelope-enhancement studies was the high level of speech perception of participants in this study (mean > 95% correct for 60 dBA speech in quiet). For example, the average speech score with ACE in the study of

Azadpour and Smith (2016) was only 40%, and a significant benefit was shown with envelope-enhancement. Looking at the individual results for speech in noise in this study, subjects with the lowest average scores (S1, S4, S5, and S7) also showed the most benefit with increased RM ratios. Subjects with lower speech understanding may benefit more from enhanced representations of the temporal envelope, while subjects with good speech understanding might not require such high modulation depths to understand speech.

V. CONCLUSION

A speech processing strategy that encoded temporal envelopes using simultaneous AM and RM, ARTmod, was tested using 60 dBA clean speech, 60 dBA speech in babble noise (+10 dB SNR), and 40 dBA clean speech. The amount of RM was varied to determine whether RM could constructively combine with AM to enhance the perception of temporal envelopes, and therefore improve speech perception. Speech scores in quiet and in noise at 60 dBA significantly improved with increased RM ratios, with the highest average scores at RM ratios of 4 and 10. At the lowest RM ratio, average speech scores were poorest for all speech conditions, compared to stimuli with higher RM ratios. These results indicate that the addition of RM effectively enhanced the perception of temporal speech envelopes.

VI. ACKNOWLEDGEMENTS

We are grateful to our dedicated participants, who graciously volunteered time and energy in many test sessions. TB is sponsored by the Melbourne International Research Scholarship and the Melbourne International Fee Remission Scholarship. The research was supported by a Veski Fellowship to CMM. The Bionics Institute acknowledges support it receives from the Victorian Government through its Operational Infrastructure Support Program.

See supplementary material at [URL will be inserted by AIP] for electrodograms for each RM amount for 40 dBA clean speech and 60 dBA speech in +10 dB SNR babble noise, and for the stimulation activity histogram for the Weatherby *et al.* (2003) data.

REFERENCES

Arai, T., and Greenberg, S. (**1998**). "Speech intelligibility in the presence of cross-channel spectral asynchrony", in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on* edited by 933-936.

Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W.-D., and Hamzavi, J.-S. (**2007**). "Speech and music perception with the new fine structure speech coding strategy: preliminary results," Acta oto-laryngologica **127**, 1298-1303.

Azadpour, M., and McKay, C. M. (**2014**). "Processing of speech temporal and spectral information by users of auditory brainstem implants and cochlear implants," Ear and hearing **35**, e192-e203.

Azadpour, M., and Smith, R. L. (**2016**). "Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing," Hearing research **342**, 48-57.

Brochier, T., McDermott, H. J., and McKay, C. M. (**2017**). "The effect of presentation level and stimulation rate on speech perception and modulation detection for cochlear implant users," The Journal of the Acoustical Society of America **141**, 4097-4105.

Brochier, T., McKay, C., and McDermott, H. (**2018**). "Rate modulation detection thresholds for cochlear implant users," The Journal of the Acoustical Society of America **143**, 1214-1222.

De Ruiter, A. M., Debruyne, J. A., Chenault, M. N., Francart, T., and Brokx, J. P. (**2015**). "Amplitude Modulation Detection and Speech Recognition in Late-Implanted Prelingually and Postlingually Deafened Cochlear Implant Users," Ear and hearing **36**, 557-566.

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," Journal of Speech, Language, and Hearing Research **40**, 1201-1215.

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," The Journal of the Acoustical Society of America **110**, 1150-1163.

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

Fu, Q.-J., and Galvin III, J. J. (**2001**). "Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users," The Journal of the Acoustical Society of America **109**, 1166-1172.

Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," Neuroreport **13**, 1635-1639.

Garadat, S. N., Zwolan, T. A., and Pfingst, B. E. (**2012**). "Across-site patterns of modulation detection: Relation to speech recognitiona)," The Journal of the Acoustical Society of America **131**, 4030-4041.

Garadat, S. N., Zwolan, T. A., and Pfingst, B. E. (**2013**). "Using temporal modulation sensitivity to select stimulation sites for processor MAPs in cochlear implant listeners," Audiology and Neurotology **18**, 247-260.

Geurts, L., and Wouters, J. (**1999**). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," The Journal of the Acoustical Society of America **105**, 2476-2484.

Gnansia, D., Lazard, D. S., Léger, A. C., Fugain, C., Lancelin, D., Meyer, B., and Lorenzi, C. (**2014**). "Role of slow temporal modulations in speech identification for cochlear implant users," International journal of audiology **53**, 48-54.

Holden, L. K., Vandali, A. E., Skinner, M. W., Fourakis, M. S., and Holden, T. A. (**2005**). "Speech recognition with the advanced combination encoder and transient emphasis spectral maxima strategies in nucleus 24 recipients," Journal of Speech, Language, and Hearing Research **48**, 681-701.

Koning, R., and Wouters, J. (**2016**). "Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users," Hearing research **342**, 13-22.

Luo, X., and Fu, Q.-J. (**2007**). "Frequency modulation detection with simultaneous amplitude modulation by cochlear implant users," The Journal of the Acoustical Society of America **122**, 1046-1054.

Magnusson, L. (**2011**). "Comparison of the fine structure processing (FSP) strategy and the CIS strategy used in the MED-EL cochlear implant system: speech intelligibility and music sound quality," International Journal of Audiology **50**, 279-287.

McKay, C. M., and McDermott, H. J. (**1998**). "Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals," The Journal of the Acoustical Society of America **104**, 1061-1074.

McKay, C. M., Remine, M. D., and McDermott, H. J. (**2001**). "Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation," The Journal of the Acoustical Society of America **110**, 1514-1524.

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (**2003**). "A practical method of predicting the loudness of complex electrical stimuli," The Journal of the Acoustical Society of America **113**, 2054-2063.

McKay, C. M., Lim, H. H., and Lenarz, T. (**2013**). "Temporal Processing in the Auditory System," Journal of the Association for Research in Otolaryngology **14**, 103-124.

Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Transactions on Biomedical Engineering **52**, 64-73.

Nie, K., Barco, A., and Zeng, F.-G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear and hearing **27**, 208-217.

Oxenham, A. J., and Moore, B. C. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hearing research **80**, 105-118.

Oxenham, A. J. (**2001**). "Forward masking: Adaptation or integration?," The Journal of the Acoustical Society of America **109**, 732-741.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). "Linear and nonlinear processes in temporal masking," Acta acustica united with acustica **88**, 348-358.

Saberi, K., and Perrott, D. R. (**1999**). "Cognitive restoration of reversed speech," Nature **398**, 760.

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

Shannon, R. V. (**2002**). "The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design," American Journal of Audiology **11**, 124-127.

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 Nucleus 24 cochlear implant system," Ear and hearing **21**, 608-624.

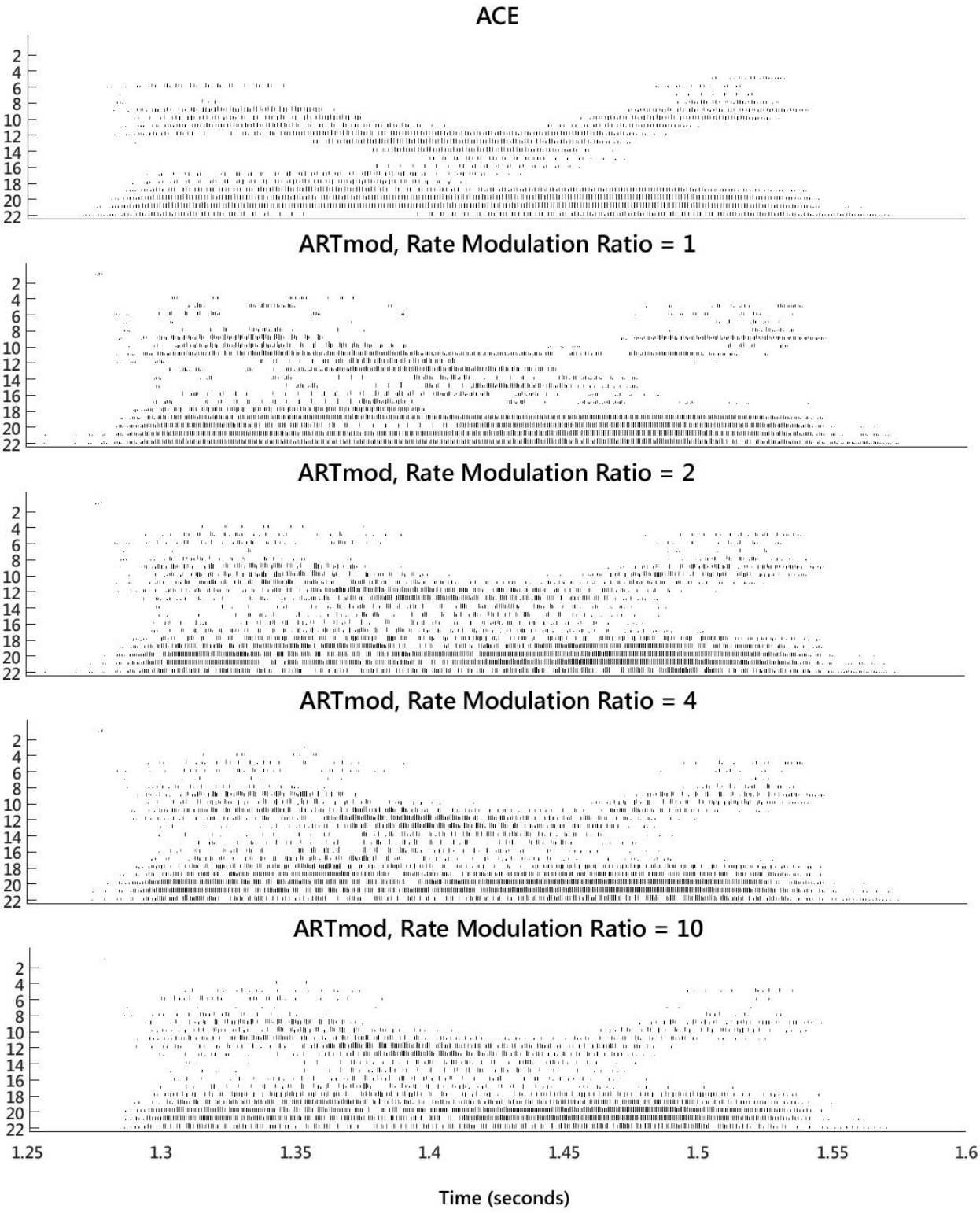
Vandali, A. E. (**2001**). "Emphasis of short-duration acoustic speech cues for cochlear implant users," The Journal of the Acoustical Society of America **109**, 2049-2061.

Weatherby, A., Henshall, K. R., and McKay, C. M. (**2003**). "The effects of frequency response on speech perception for cochlear implant users," Journal of the American Academy of Audiology **14**, 582-591.

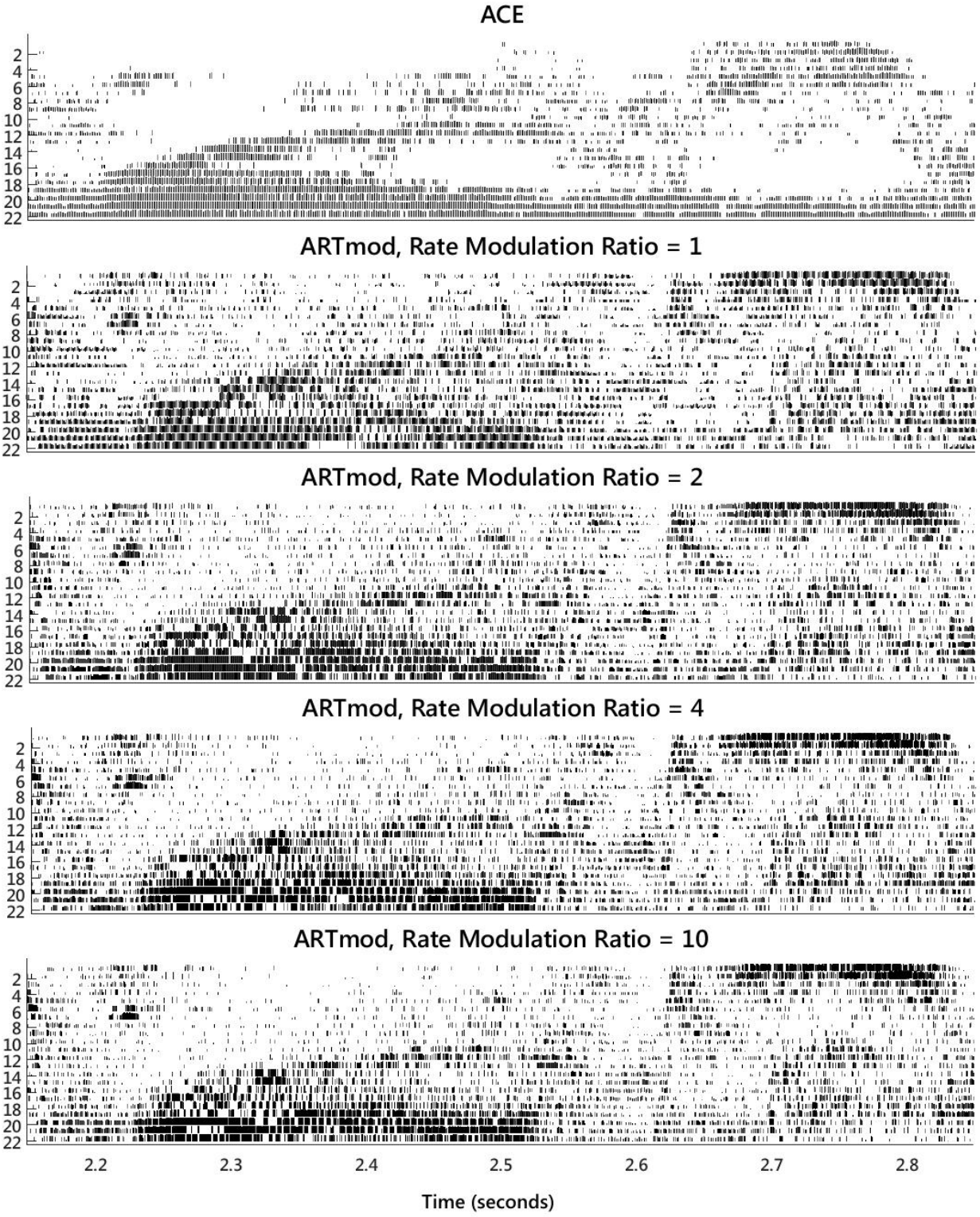
Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., and Rubinstein, J. T. (**2011**). "Acoustic temporal modulation detection and speech perception in cochlear implant listenersa)," The Journal of the Acoustical Society of America **130**, 376-388.

Wouters, J., McDermott, H. J., and Francart, T. (**2015**). "Sound Coding in Cochlear Implants: From electric pulses to hearing," Signal Processing Magazine, IEEE **32**, 67-80.

Supplementary Material



SuppPub1. Electrodograms displaying the sequence of pulse amplitudes over time at each active electrode. Each stimulation sequence represents the word 'VERY' spoken at a presentation level of 40 dBA. The first panel was encoded using the ACE strategy, and panels 2 through 5 were encoded using the ARTmod strategy with an RM ratio of 1, 2, 4, and 10 respectively.

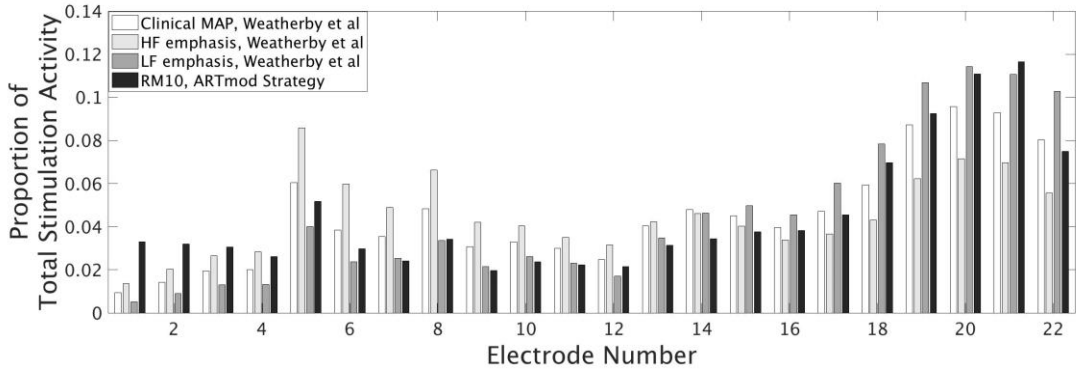


573

574 **SuppPub2. Electrodograms displaying the sequence of pulse amplitudes over time at each**

575 **active electrode. Each stimulation sequence represents the word 'WORDS' spoken at a**

presentation level of 60 dBA in +10 dB SNR speech babble noise. The first panel was encoded using the ACE strategy, and panels 2 through 5 were encoded using the ARTmod strategy with an RM ratio of 1, 2, 4, and 10 respectively.



SuppPub3. Envelope magnitude histogram for 60 dBA speech, displaying the average proportion of speech stimuli represented by each electrode for the different emphasis filters in Weatherby *et al.* (2003) and the RM10 condition in the present study. Transformed envelope magnitude values were summed across 16 sentence stimuli for each individual electrode, and then those 22 values were divided by the total sum across channels.