5-2014

An Evaluation of an Auditory Neurophysiological Model

Alysha Nicole Waters
Utah State University

AN EVALUATION OF AN AUDITORY NEUROPHYSIOLOGICAL MODEL

by

Alysha Nicole Waters

Thesis submitted in partial fulfillment

of the requirements for the degree

of

HONORS IN UNIVERSITY STUDIES

WITH DEPARTMENTAL HONORS

In

Biology

in the Department of Biology

Approved:

Thesis/Project Advisor

Dr. Donal G. Sinex

Departmental Honors Advisor
Thesis Committee Member
Dr. Brett Adams

Director of Honors Program

Dr. Nicholas Morrison

UTAH STATE UNIVERSITY

Logan, UT

Spring 2014

ABSTRACT

Individuals with normal hearing are adept at understanding speech in the presence of noise, such as other speakers or environmental sounds. In contrast, individuals with hearing loss struggle to understand speech in the same adverse conditions. Neural processing in the inferior colliculus (IC) of the brainstem appears to contribute to the ability to separate simultaneous competing sounds. A computational model developed in the Sinex lab reproduces the responses of IC neurons to complex sound mixtures. It seems likely that the model can be applied to improve the processing of speech in noise. The computational model's effectiveness at improving the processing of speech in noise is evaluated through a perceptual experiment which uses the model to process sentences that are then presented to listeners. The experiment's data are analyzed to evaluate the pattern of errors. The analysis shows that low frequency speech features are being accurately transmitted by the model while high frequency speech features are not. This pattern suggests ways in which the computational model may be improved. Possible technological and clinical applications of the computational model for individuals with hearing loss will also be discussed.

CONTENTS

FIGURES AND TABLES

Tables

Figures

1. Introduction

Individuals with normal hearing are adept at understanding speech in the presence of noise, such as other speakers or environmental sounds. In contrast, individuals with hearing loss struggle to understand speech in the same adverse conditions. Individuals with normal hearing are able to take in several sounds at once and process the sounds individually, a process called sound-source determination (Yost, 1992). Sounds from a listening environment sum together and enter the auditory system in a single waveform (Yost, 1992). Sounds have different frequency, time, and amplitude components (Yost, 1992). The auditory system must then match the components of frequency, time, and amplitude that belong together from a single source (Yost, 1992).

Auditory neurophysiological models are better able to replicate the auditory pathway by breaking the auditory system into smaller components which mimic parts of a normal listener's auditory system. When all of the smaller components are added together, they provide an overview of how the auditory system works. Haykin and Chen (2005) comment that a computational model of the auditory system does not require that every part of the auditory system be replicated in a model, just the parts that provide enough information to form a picture of how the auditory system works. Researchers can contribute a more accurate component to a computational model; however a final computational model that describes how the auditory system works will depend on the combined efforts of the hearing research community (Meddis & Lopez-Poveda, 2010b).

Computational models of the auditory system have both technological and clinical applications. Computational models might be used to improve voice recognition programs in computers (Meddis & Lopez-Poveda, 2010a) such as HeyTell and Siri. Computational methods which aim to reduce the noise in speech can also be applied to hearing aids and cochlear implants for those who struggle with hearing loss (Bentler and Chiou, 2006; Wang, 2008; Wang et al., 2009).

The neurological mechanisms responsible for sound-source determination are just beginning to be understood; for example Sinex et al. (2005) have researched responses of the brainstem's inferior colliculus (IC) neurons in chinchillas. The results from their study have produced a computational model for sound source determination in the auditory system applicable to humans (Sinex et. al 2005, Sinex, 2008). The model developed by the Sinex lab imitates the discharge responses of neurons in the IC thought to be responsible for sound source determination (Sinex et al., 2005).

As a first step toward evaluations, the model was used to process speech consonants and vowels. This experiment focused on the ability of the processing method to transmit three speech features: consonant voicing, consonant place of articulation, and vowel place of articulation. The processed speech sounds were then presented to listeners in a psychophysical experiment. Evaluation of the model's ability to reduce noise was accomplished by calculating Information Transfer (IT).

2. Computational Phase

2.1 Model Description

The auditory neurophysiological model consists of four sequential stages and a neurophysiologically-based binary mask (NBBM). The four stages simulate the responses of neurons in the auditory nervous system. The NBBM is the part of the model that attempts to reduce the noise in speech.

2.1.1 Stage 1

Stage 1 of the model is provided by Zilany et al. who have made their computer program available for other auditory researchers. Stage 1 simulates the synapse responses of the inner hair cell (IHC) and auditory nerve (AN) fibers. Zilany et al. (2009) have been able to more closely match physiological data at the IHC-AN synapse, the first in the auditory pathway, by including power-law adaptation in their model to explain the long term adaptation at the IHC-AN synapse.

2.1.2 Stage 2

Stage 2 of the model simulates the responses of cochlear nucleus neurons (Sinex, 2008). Cochlear nucleus neurons, located in the brainstem, receive their input from AN fibers. Cochlear neurons relay information from the AN fibers to inferior colliculus (IC) neurons. Neural discharge responses from cochlear nucleus neurons can resemble those of AN fiber discharge responses, IC discharge responses, or responses in between AN fibers and IC neurons (Sinex, 2008). The computer program for this stage was developed in the Sinex lab.

2.1.3 Stage 3

Stage 3 of the model simulates the responses of IC neurons located in the midbrain. IC neurons are important to the auditory nervous system because they are believed to make a major contribution to sound source determination (Sinex, 2005). IC neurons have been shown to exhibit enormous responses to mistuned stimuli in comparison to AN fibers, making the inclusion of IC neurons in auditory neurophysiological models essential (Sinex, 2005). Additionally, neural inhibition has been found to play an important role in determining the temporal discharge patterns of IC responses (Li et al., 2006). The computer program for this stage of the model was developed in the Sinex lab.

2.1.4 Stage 4

Stage 4 of the model is the last in a sequence of neural responses for the auditory nervous system. In this stage, the model simulates the responses of modulation sensitive neurons in the auditory pathway. The auditory cortex contains regions of neurons that are excitatory and regions of neurons that are inhibitory (Schreiner, 1995). Excitatory regions occur in the ventral auditory primary cortex and between the dorsal and central auditory primary cortex, while inhibition occurs at the dorsal side of the primary and secondary auditory cortex (Schreiner, 1995). Many of the neurons in the auditory pathway exhibit selectivity for modulation frequency. Neurons with this selectivity will have a stronger response to the amplitude envelopes of sounds that rise and fall at particular rate (Sinex et al., 2003). The simulated neuron response program was developed in the Sinex lab.

2.1.5 Neurophysiologically-based Binary Mask (NBBM)

At the end of Stage 4 the NBBM is applied in order to extract the target speech of the sentence from the background noise. The objective of the NBBM is to retrieve the entire sentence noise-free.

The NBBM is a modification of an "ideal binary mask." An ideal binary mask has the ability to reduce the noise in speech; however prior knowledge of the speech is required (Wang, 2005). This is impractical for everyday use in a hearing device because conversations are not scripted. NBBMs are a unique and novel approach because they rely on the physiological responses of the auditory system.

For a target signal mixed with some type of interference signal, the binary mask acts as a filter (Brungart et al., 2006). If the interference signal is stronger than the target signal, that part of the overall signal is designated as a zero (Brungart et al., 2006). Conversely, a part of the signal where the target signal is greater than the interference signal is designated as with a one (Brungart et al., 2006). The binary mask rejects the parts of the signal designated with a zero and passes the parts of the signal designated with a one (Brungart et al., 2006).

2.2 Stimulus Processing

2.2.1 Speech Features

Consonant speech features commonly studied in auditory psychophysical experiments are voicing, nasality, affrication, duration, place of articulation, and envelope (Miller & Nicely, 1955; Sagi & Svirsky, 2008). For simplicity, two consonant speech features were used in this experiment; those of consonant voicing and consonant place of articulation.

A consonant is considered voiced if the vocal cords vibrate during a consonant's articulation, while a consonant is considered voiceless if the vocal cords do not vibrate during articulation (Miller & Nicely, 1955). The consonants b, d, and g are examples of voiced consonants and were used in this experiment. The consonants p, t, and k are examples of voiceless consonants used in this experiment.

Consonant place of articulation describes consonants according to where the vocal tract is constricted to produce the consonant. The consonants p and b are bilabial, referring to a constriction formed at the lips (Miller & Nicely, 1955). Consonants t and d are alveolar; they are articulated with a constriction behind the teeth (Miller & Nicely, 1955). Consonants k and g are velar; they are articulated with a constriction between the tongue and the velum of the soft palate (Miller & Nicely, 1955; Wells & House, 1995).

The third speech feature used was vowel place of articulation. Vowel place of articulation can be divided into dimensions of tongue height, tongue position, and lip posture (Pfitzinger & Niebuhr, 2011). For this experiment, vowel place of articulation refers to the position of the tongue for the phonemes /ae/ (as in "hat") and /ah/ (as in "hot"). Tongue position has significance to vowels because it determines the formants, or peaks in the spectral profile of a sentence (see Sinex, 2012 for a review). Each vowel phoneme has a characteristic difference in peaks, or formants (Sinex, 2012). Using the formant information provided by tongue position, listeners are able to distinguish between different vowel phonemes (Sinex, 2012).

2.2.2 Materials and Methods

All sentences were prerecorded with a microphone (Shure Beta 58A) in a sound booth (Industrial Acoustics Company, New York) and digitized using custom software written in Matlab (The Mathworks, Natick MA). Each sentence contained the carrier phrase "Say the word" and one of twelve consonant-vowel syllables (e.g., /bah/ as in "bother" or bae as in "bat"). All consonant-vowel syllables contained one of six consonants (b, d, g, p, t, k) and one of two vowel phonemes (/ae/ or /ah/). The carrier phrase was recorded first. The twelve consonant vowel syllables were later recorded and added to the end of the carrier phrase using Adobe Audition 1.5. The sentences were mixed with noise and processed through the model using Matlab.

2.3 Model Results

The neurophysiological model of the auditory system shows reduction of noise in speech. The figure below (Fig. 1) is a spectrogram of the sentence "Say the word /tah/" at -4 dB SNR. The SNR level of -4 is approximate to the level that the psychophysical data were collected. The sentence "Say the word /tah/" is representative; all sentences show the same general results. The vertical axis of Fig. 1 is the frequency in kHz and the horizontal axis is the time in msec. The color maps on the right show the amplitude measured in dB sound pressure level (dB SPL). The sound pressure level at 60 dB SPL is 60 dB above 20μPa. The range of 40 to 60 dB SPL was chosen to highlight the amplitude peaks and valleys. The red coloring in the spectrogram indicates spectral components with the highest amplitude.

The upper panel of Fig. 1 shows the spectrogram of the original recorded sentence "Say the word /tah/." There are both high and low frequency components to this sentence. The high frequency components of the sentence are located in the top portion of the upper panel. The low frequency components of speech are shown in the bottom portion of the upper panel.

The middle panel of Fig. 1 shows the sentence processed by the model, as indicated by the label "re-synthesized speech." The sentence (shown in the lower panel) has undergone simulation responses of the neurons in the model (AN, cochlear nucleus neurons, IC, and modulation sensitive neurons) and had a NBBM applied. The aim is for the processed sentence (lower panel) to be identical or nearly identical to the original sentence (upper panel).

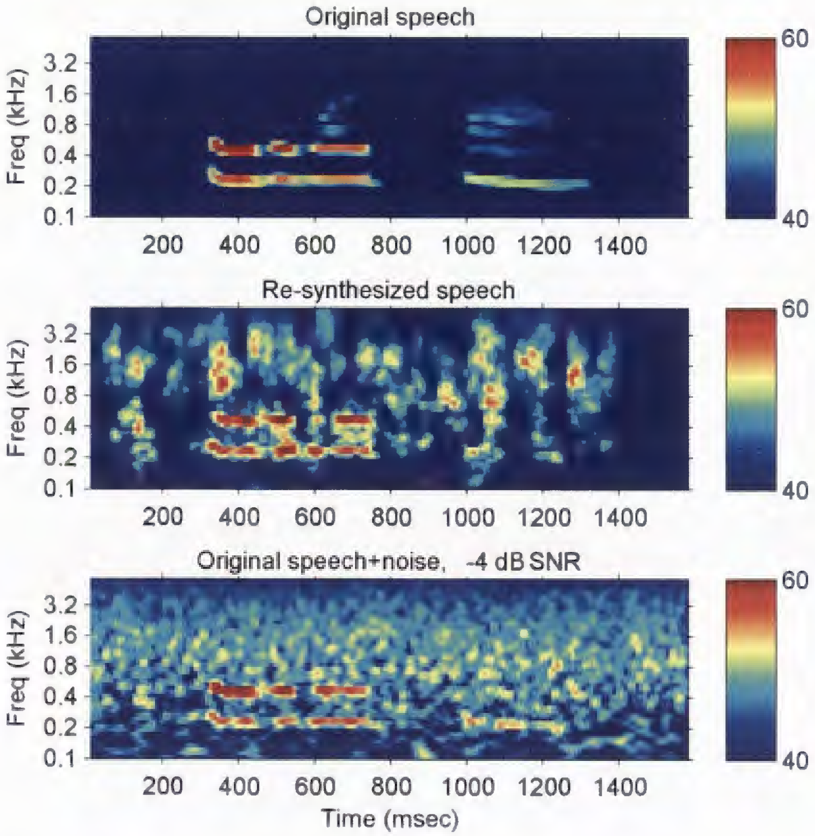


Figure 1. Spectrograms of the sentence "Say the word /tah/" at -4 dB SNR. Frequency is provided in kHz on the left vertical axis. Time is shown in msec on the horizontal axis. The color maps on the right measure amplitude intensity in dB SPL. Upper panel: Spectrogram of the original sentence. Middle panel: Spectrogram of the sentence processed by the model. Lower panel: Spectrogram of the sentence mixed with noise. Noise appears as a rough overlay.

The lower panel of Fig. 1 shows the original sentence mixed with noise. The noise disruption of the spectral pattern to the original sentence (upper panel) is clearly visible, as light blue and yellow patches. The high frequency speech features of the lower panel are nearly indistinguishable; they have almost completely been masked with noise. The low frequency speech features are also masked with noise, however not to the extent of high frequency speech features.

The processed sentence (lower panel) is presented to the listener during the perceptual phase of the experiment. Figure 1 depicts the efficiency of the computational model to reduce noise.

The lower panel shows a loss in the high frequency components of speech when compared to the original sentence in the upper panel. However, the model does succeed in the main objective: reducing noise in speech. The sentence processed by the model (middle panel) shows clear reduction in noise when compared to the sentence mixed with noise (lower panel).

3. Psychophysical Phase

3.1 Listeners

Six listeners (2 males, 4 females; median age 24.5 years; age range of 20-63 years) participated in this experiment. All procedures were approved by the Institutional Review Board at Utah State University. Four listeners, including the author, were volunteers. The other two listeners were provided an hourly compensation for their time. Listeners provided Informed Consent and were assigned a personal identification number for confidentiality purposes. All listeners were given a brief hearing test prior to starting the experiment. Only subjects with normal hearing thresholds were allowed to participate. After completing the hearing test, subjects listened in the sound booth to a series of sentences.

3.2 Materials

All sentences were played over Sennheiser HD 280 pro 64 Ω headphones. Custom Matlab software was used to present the hearing test prior to the experiment, as well as the re-synthesized sentences to listeners during the experiment. The stimulus levels were specified in decibel signal to noise ratio (dB SNR). For this experiment dB SNR is defined as the speech level of the sentences in dB minus the noise level in dB. Listeners hearing the sentence mixed with noise are expected to have thresholds at -6 dB SNR when unprocessed by the model (Sinex, Unpublished Observations).

Listeners' data were analyzed using custom Matlab software, GraphPad Prism 4, and Microsoft Excel. Results from each listener were pooled based on the three speech features used in this study and statistical analysis was conducted with $\alpha = 0.05$ using a one way repeated-measures analysis of variance (RMANOVA), also using Graphpad Prism.

3.3 Methods

The parts of the target speech, isolated by the NBBM of the model, were presented to the listeners in the experiment. All sentences were repeated four times throughout one block, for a total of 48 sentences during one block. Listeners continued to the next block of sentences, decreasing SNR by 2 dB each time. When the SNR range that produces a proportion correct or p(c), of 0.50 was identified. The listener then completed several blocks at that SNR range. A p(c) of 0.5 indicated a listener had half correct, half incorrect responses to the stimuli, ensuring errors to test the model's effectiveness. After closer evaluation, the p(c) over these blocks ranged from 0.30 to 0.70. The listener's data were coded into matrices with a stimulus and the possible listener responses to that stimulus, known as a confusion matrix, to analyze the pattern of errors (Miller & Nicely, 1955).

3.4 Results

3.4.1 Confusion Matrices

The confusion matrix of all listeners' data in this experiment is shown in Table 1. The stimuli presented to the listener appear on the vertical axis while the listeners' responses appear on the horizontal axis. Numbers on the diagonal axis of the confusion matrix show a correct response to a stimulus. Numbers off the diagonal axis that are greater than zero show the listeners' incorrect responses to the presented stimulus. As shown in Table I, the pattern of errors in the perception of speech is not random. Some errors are very likely to occur (e.g., the stimulus /tah/ confused as the stimulus /pah/) while other errors are not likely to occur at all (e.g., the stimulus /tah/ confused as the stimulus /bae/).

TABLE I. Confusion matrix of all listeners' data. The vertical axis corresponds to the stimulus that was presented (e.g., /bah/ indicates the sentence "Say the word /bah/" was played). The horizontal axis corresponds to the listener's response to the presented stimulus (e.g., /pah/ indicates the listener choose the box labeled '/pah/'). The confusion matrix shows the pattern of errors from all listeners' data. The numbers in the blue diagonal axis indicate the amount of correct pairings between a stimulus and the listener's response. Numbers greater than zero off the blue diagonal axis indicate a listener's response to a presented stimulus was incorrect.

Listener's Response

Presented Stimuli	/bah/	/dah/	/gah/	/pah/	/tah/	/kah/	/bae/	/dae/	/gae/	/pae/	/tae/	/kae/
/bah/	88	47	18	53	96	89	0	1	1	1	2	0
/dah/	1	329	57	0	1	0	0	8	0	0	0	0
/gah/	2	51	323	2	0	2	0	13	3	0	0	0
/pah/	39	5	3	257	72	15	0	0	0	3	1	1
/tah/	45	1	3	225	74	39	0	0	0	6	1	2
/kah/	108	5	38	112	48	72	9	0	0	1	1	2
/bae/	6	2	0	0	0	1	296	69	14	8	0	0
/dae/	3	12	1	0	1	0	74	278	26	0	1	0
/gae/	0	0	19	0	0	0	0	46	330	0	0	1
/pae/	0	0	0	35	19	0	2	1	0	199	58	82
/tae/	1	0	0	4	11	0	2	2	3	149	203	21
/kae/	0	0	0	5	1	4	26	2	2	105	23	228

Table I can be collapsed into a series of smaller matrices based upon the three binary speech features studied in this experiment. Tables II, III, and IV are the confusion matrices of the individual speech features in the consonant-vowel syllables presented to listeners. Like Table I, the numbers in the diagonal axis represent a correct pairing of stimulus and response while the numbers greater than zero off the diagonal axis show incorrect pairing of stimulus and response. The column totals in Tables II, III, and IV represent the number of times that a stimuli was presented. The row totals indicate the number of times a specific speech feature was choosen. If there were no perceptual errors, the column totals for the presented stimuli and the row totals of the listener's response would be the same.

TABLE II. The confusion matrix shows the pattern of errors for the speech feature of consonant voicing. The vertical axis corresponds to the stimuli presented based upon a consonant's voicing. The horizontal axis corresponds to the listener's response based upon a consonant's voicing. Correct pairing of a stimulus with the consonant's voicing is shown by the numbers on the diagonal axis, while incorrect pairings are off the diagonal axis. A presented voiced consonant was confused to be a voiceless consonant 258 times. The row of totals corresponds to the number of times a listener chose voiced or voiceless. The column of totals corresponds to the number of times a voiced or voiceless consonant was presented.

		Listeners' Response		
		Voiced	Voiceless	Totals
Presented Stimuli	Voiced	2118	258	2376
	Voiceless	297	2079	2376
	Totals	2415	2337	

TABLE III. The confusion matrix shows the pattern of errors for the speech feature of consonant place of articulation. The vertical axis corresponds to the stimuli presented based upon a consonant's place of articulation. The horizontal axis corresponds to the listener's response based upon a consonant's place of articulation. Correct pairing of a stimulus with the consonant's place of articulation is shown by the numbers on the diagonal axis, while incorrect pairings are off the diagonal axis. A consonant articulated bilabially was confused to be a consonant articulated at the alveolars 373 times. The row of totals corresponds to the number of times a listener chose a place of articulation. The column of totals corresponds to the number of times a consonant's place of articulation was presented.

		Listeners' Response			
		Bilabial	Alveolar	Velar	Totals
Presented Stimuli	Bilabial	987	373	224	1584
	Alveolar	510	922	152	1584
	Velar	370	190	1024	1584
	Totals	1867	1485	1400	

TABLE IV. The confusion matrix shows the pattern of errors for the speech feature of vowel place of articulation. The two vowel phonemes /ae/ and /ah/ differ in their place of articulation. The vertical axis corresponds to the stimuli presented based upon a vowel's place of articulation. The horizontal axis corresponds to the listener's response based upon a vowel's place of articulation. Correct pairing of a stimulus with the vowel's place of articulation is shown by the numbers on the diagonal axis, while incorrect pairings are off the diagonal axis. The presented vowel phoneme /ae/was confused with the vowel phoneme /ah/ 56 times in 2376 trials. The row of totals corresponds to the number of times a listener chose a vowel phoneme. The column of totals corresponds to the number of times a vowel phoneme of articulation was presented.

		Listeners' Response		
		/ae/	/ah/	Totals
Presented Stimuli	/ae/	2118	258	2376
	/ah/	297	2079	2376
	Totals	2445	2307	

3.4.2 IT Analysis

The confusion matrices from this experiment show perceptual errors; however it is not possible to quantify these errors from a matrix. An equation provided by Miller and Nicely (1955) has the ability to quantify errors using a measure of IT. The equation for IT is shown below:

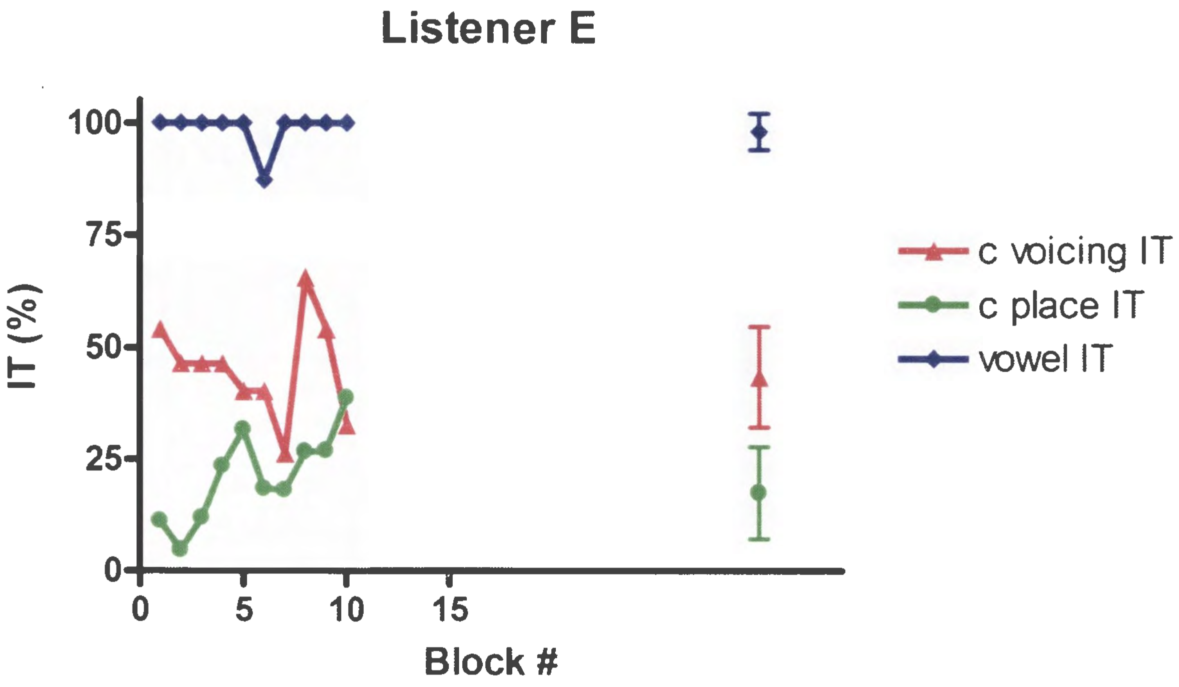
$$\mathrm{IT} = \frac{I}{H}$$

$$= \frac{-\sum_{x} \frac{n_x}{N} \log \frac{n_x}{N} - \sum_{y} \frac{n_y}{N} \log \frac{n_y}{N} + \sum_{x} \sum_{y} \frac{n_{xy}}{N} \log \frac{n_{xy}}{N}}{-\sum_{x} p_x \log p_x}.$$

(1)

In Equation (1) IT is the measure of information transmitted. The amount of information in the listeners' responses, calculated from the confusion matrix, is represented by the variable I. The variable H represents the amount of information in the stimuli. The total number of sentences presented to a listener is indicated by N. Variable x corresponds to stimuli presented while variable y corresponds to the listeners' responses. Variable n_x is therefore the number of times a sentence was presented (the sum of a row in a confusion matrix) and n_y is the number of times a listener made a response (the sum of a column in a confusion matrix). The variable p_x corresponds to the true probability that a certain consonant vowel syllable was presented to the listener. For a more detailed description of this equation refer to Sagi and Svirsky (2008).

Data from one representative listener (Listener E) at -4 dB SNR are shown in Figure 3. The vertical axis shows IT in percent for each speech feature. The horizontal axis of Block # indicates how many blocks of sentences the individual listened to with a p(c) close to 0.50. The lone symbols apart from the rest of the data represent the average IT of a speech feature for an individual.



In Fig. 3, vowel IT refers to the IT for vowel place of articulation, c place IT refers to the IT value for consonant place of articulation, and c voicing IT refers to the IT for consonant voicing. The speech feature vowel place of articulation was most accurately transmitted by the model with an average IT of 98.74%. This means that over 98% of the information for the vowel place of articulation was transmitted by the model to the listener. With an average of 21.16% IT, consonant place was the least accurately transmitted speech feature by the model. Averaging 45.15 % IT, the speech feature of consonant voicing falls between vowel place of articulation and consonant place of articulation. Results from all other listeners are shown in Figures 3-7.

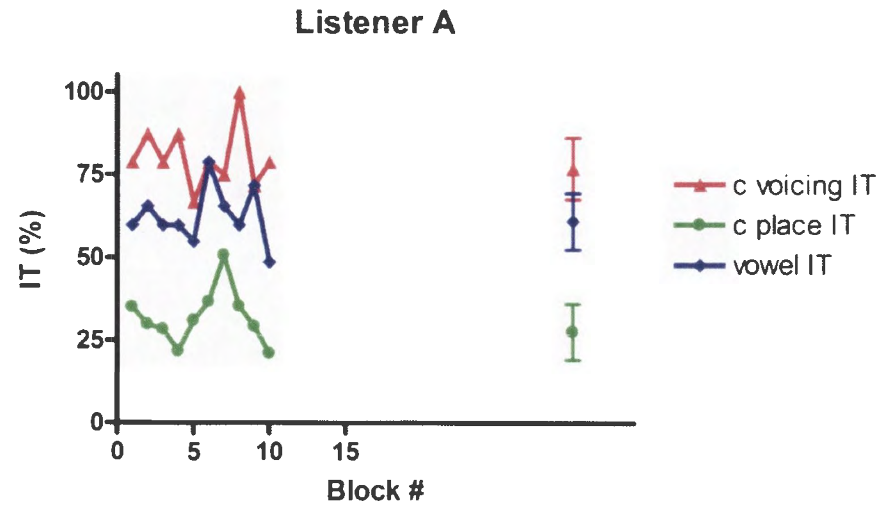


Figure 3. Results from listener A.

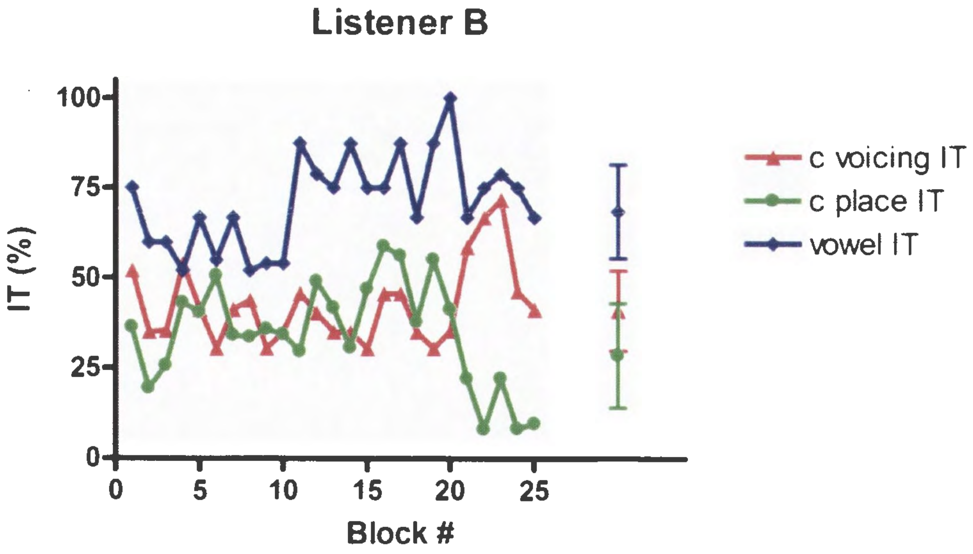


Figure 4. Results from listener B.

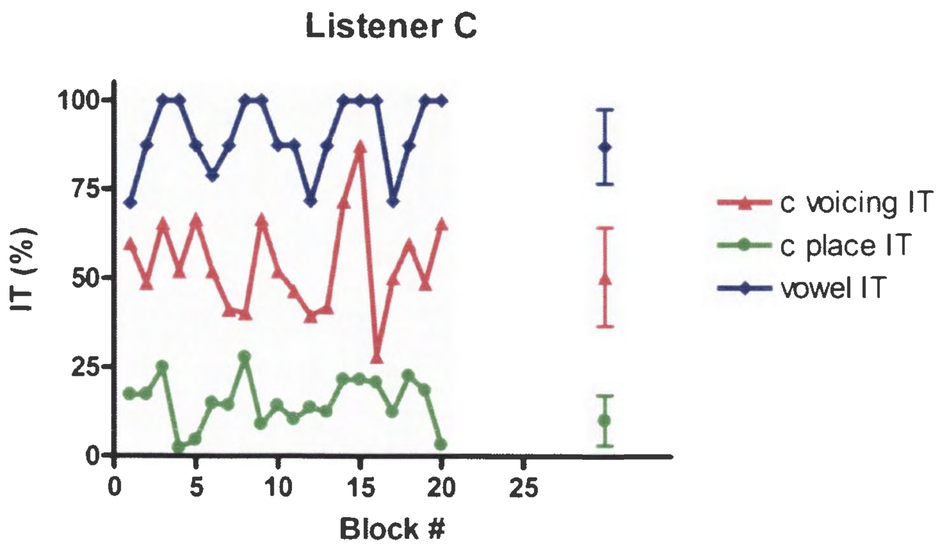


Figure 5. Results from listener C.

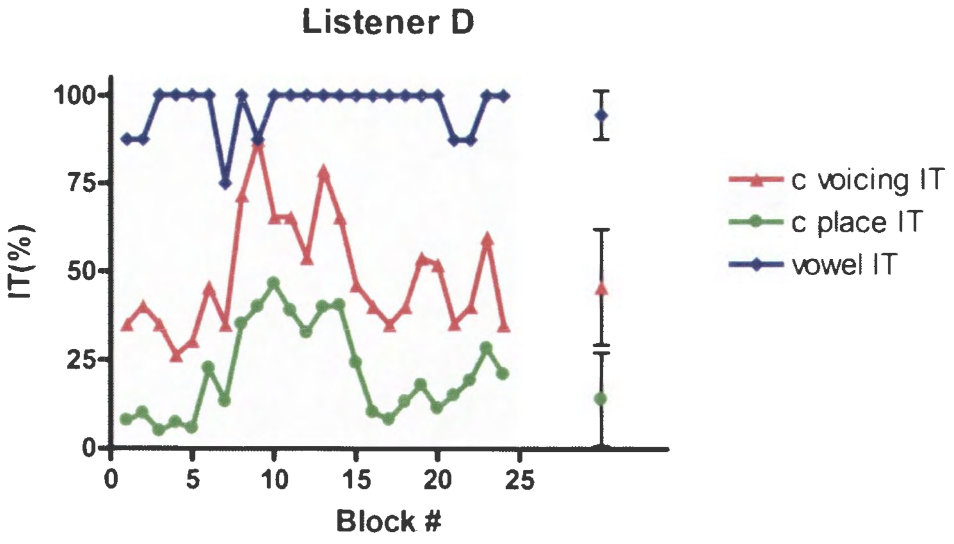


Figure 6. Results from listener D.

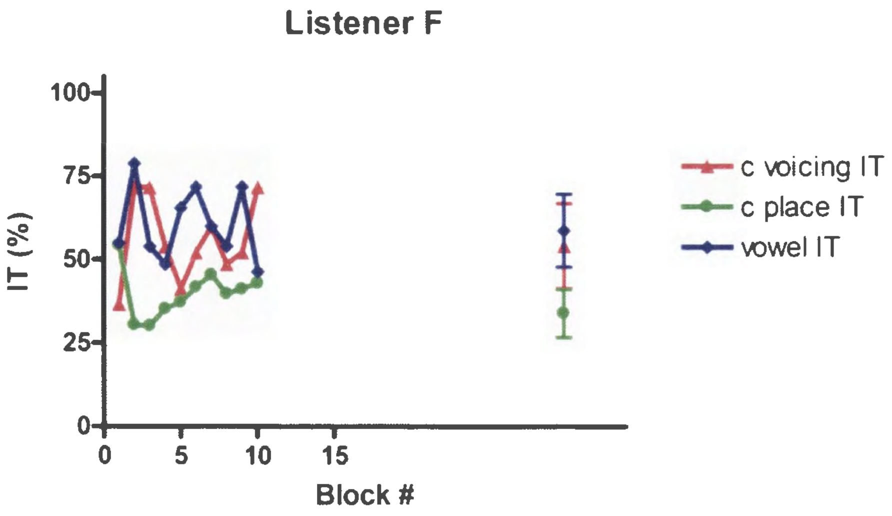


Figure 7. Results from listener F.

Figure 8 displays the pooled data for the listeners in this experiment. The bar on the far right indicates a vowel place of articulation with an IT value of 77.02 %. Vowel place of articulation was most accurately transmitted by the model. The middle bar representing consonant place of articulation has an IT value of 17.57 % and was least accurately transmitted by the model. The bar on the far left represents the consonant voicing speech feature. With an IT value of 48.02% this speech feature falls between the other speech features of this experiment. Statistical analysis using RMANOVA showed the means of consonant voicing, consonant place of articulation, and vowel place of articulation significantly different ($p < 0.0005$).

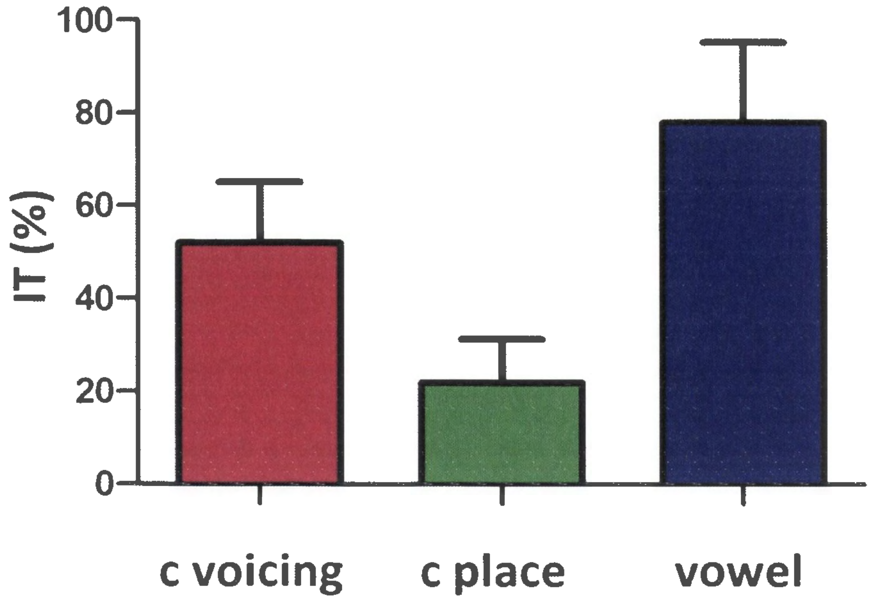


Figure 8. Results from all listeners. The horizontal axis designates the speech features. The vertical axis gives the IT value in percent. The blue bar (right) represents the pooled IT value for vowel place of articulation. The green par (middle) represents the pooled IT value for consonant place of articulation. The pink bar (left) represents the pooled IT value for consonant voicing.

4. Discussion

Auditory computational models are currently being developed to reduce the noise in speech (Sinex et al., 2005; Zilany et al., 2009). Research advances in noise rejection could be implemented in speech recognition programs for computers (Meddis & Lopez-Poveda, 2010a). Noise reduction in speech recognition programs could lead to decreased errors caused by interfering noise in speech. A noise reduction program could also enable people to use speech recognition programs in an environment that presents adverse listening conditions.

Noise rejection models can also be used to improve cochlear implants and hearing devices. Those who suffer from hearing loss have a greater struggle in reducing the noise in speech compared to a normal hearing listener. The use of a hearing device is sometimes helpful, in quiet conditions; however they are not so effective in noise. Current hearing devices are met with certain problems: 1) an increase in gain limits the audibility of speech because noise is amplified along with the speech, and 2) a decrease in gain would mean a decrease in noise along with speech, defeating the purpose of a hearing device to aid those with hearing loss (Bentler & Chiou, 2006). Many that are deaf or hard of hearing find the amplified noise in a hearing device to be a problem. Some chose to not use hearing devices because of the increased noise interruption. Research advances can be used to improve the quality of life for an individual who has suffered from hearing loss, making social events more enjoyable.

The computational model developed in the Sinex lab has the ability to reduce noise in speech. The model produced a clear reduction in noise, as shown in Fig. 1. Figure 1 also shows that the high frequency components of speech are being lost while processed in the model. Similarly, results from the perceptual experiment (Figures 2-8) indicate that the speech features of consonant voicing and vowel place of articulation, which are primarily low frequency, are more accurately transmitted by the model than consonant place of articulation, which is primarily high frequency.

Low frequency neurons are able to respond directly to low frequency tones, while high frequency neurons of the auditory system respond indirectly to low frequency tones. This difference in direct or indirect response can be attributed to the wider tuning curve of high frequency neurons which allows for summation of frequencies to occur. Although two beats may have the same frequencies, they can differ in amplitude modulation. A deeply modulated beat will have a large difference in the peaks and valleys in the wave. A shallowly modulated beat will have a small difference in the peaks and valleys of the wave. Shallow modulation may contribute to the model's decreased capability to transmit high frequency components of speech.

In the perceptual experiment conducted, listeners achieved a proportion correct (p(c)) of 0.50 at SNRs ranging 0 to -4 dB when sentences were processed by the model. This indicates noise reduction by the model, as listeners are expected to have a threshold of -6 dB without processing by the model (Sinex, Unpublished Observations).

Although the high frequencies of speech are lost in processing, the computational model developed by the Sinex lab achieves the aim of reducing noise in speech. More research should be conducted to increase the reduction the noise even further than that achieved with this model. Improvements made to the computational model developed in the Sinex lab may allow the model to become beneficial in technological and clinical applications.

Acknowledgments

Supported by National Institutes of Health grant DC010615. Special thanks to Dr. Sinex for his guidance and support of this project. Thanks are also due to Dr. Adams and Richard and Karen Waters for their comments.

References

Bentler, R., & Chiou, L. K. (2006). Digital noise reduction: an overview. *Trends in Amplification. 10* (2), 67-82.

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *Journal of the Acoustical Society of America, 120* (6), 4007-4018.

Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Computation, 17*(9), 1875-1902.

Li, H., Sabes, J.H., & Sinex, D.G. (2006). Responses of inferior colliculus neurons to SAM tones located in inhibitory response areas. Hear. Res., *220*, 116-125.

Meddis, R. & Lopez-Poveda, E. A. (2010a). Auditory periphery: From pinna to auditory nerve. In R. Meddis, E. A. Lopez-Poveda, A. N. Popper, & R. R. Fay (Eds.), *Computational models of the auditory system* (pp. 7-38). New York, NY: Springer.

Meddis, R. & Lopez-Poveda, E. A. (2010b). Overview. In R. Meddis, E. A. Lopez-Poveda, A. N. Popper, & R. R. Fay (Eds.), *Computational models of the auditory system* (pp. 1-6). New York, NY: Springer.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27* (2), 1243-1252.

Pfitzinger, H. & Niebuhr, O. (2011). Historical development of the phonetic vowel systems-The last 400 years.

Sagi, E., & Svirsky, M. A. (2008). Information transfer analysis: A first look at estimation bias. *Journal of the Acoustical Society of America*, *123* (5), 2848-2857.

Schreiner, C. E. (1995). Order and disorder in auditory cortical maps. *Current Opinion in Neurobiology*, *5*, 489-496.

Sinex, D.G. (2012). Complex vowel encoding- Vowels and consonants. In K. Tremblay & R Burkard. *Translation perspectives in auditory neuroscience: Normal aspects of hearing* (pp. 435-463).

Sinex, D. G. (2008). Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. *Hearing Research*, *238* (1-2), 39-48.

Sinex, D. G. (2005). Spectral processing and sound source determination. *International Review of Neurobiology*, *70*, 371-398.

Wang, D. L. (2005). On ideal binary mask as the computational goal of auditory scene analysis. *Speech Separation by Humans and Machines*. 181:197.

Sinex, D.G., Guzik, H., Li, H., & Henderson Sabes, J. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hearing Research*, 182(*1*), 130-139.

Wang, D.L. (2008). Time-frequency masking for speech separation and its potential for hearing aid design. *Trends in Amplification*. 12(*4*) 332-353.

Wang, D.L., Kjems, U., Pedersen, M.S., Boldt, J. B., & Lunner, T. (2009). Speech intelligibility in background noise with ideal binary time-frequency masking. *The Journal of Acoustical Society of America*, 125(*4*) 2336-2347.

Wells, J. & House, J. (1995). The sounds of the International Phonetic Alphabet. UCL, London.

Yost, W. A. (1992). Auditory perception and sound source determination. *Current Directions In Psychological Science (Wiley-Blackwell)*, *1*(6) 179-184.

Zilany, M. S. A., Carney, L. H., Bruce, I. C., & Nelson, P. C. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *Journal of the Acoustical Society of America*, *126* (5), 2390-2412.

AUTHOR'S BIOGRAPHY

Born in Logan, Utah, Alysha Waters returned to Logan in 2009 after graduation from Northridge High in Layton, Utah. During the past few years at Utah State Alysha has been studying Biology and minoring in Chemistry and Psychology. She has served as a member of the Honors Student Council and a Head Mentor for the Honors Program. She is currently working on her application to medical school.