



Published in final edited form as:

Hear Res. 2015 April ; 322: 163–170. doi:10.1016/j.heares.2014.10.008.

## Gradual adaptation to auditory frequency mismatch

Mario A. Svirsky<sup>1,2</sup>, Thomas M. Talavage<sup>3,4</sup>, Shivank Sinha<sup>5</sup>, Heidi Neuburger<sup>6</sup>, and Mahan Azadpour<sup>1</sup>

<sup>1</sup>Dept. of Otolaryngology-HNS, New York University School of Medicine, New York, NY

<sup>2</sup>Center of Neural Science, New York University, New York, NY

<sup>3</sup>ECE Dept., Purdue University, West Lafayette, IN

<sup>4</sup>BME Dept., Purdue University, West Lafayette, IN

<sup>5</sup>EE Dept., Purdue University, Indianapolis, IN

<sup>6</sup>Dept. of Otolaryngology-HNS, Indiana University School of Medicine, Indianapolis, IN

### Abstract

What is the best way to help humans adapt to a distorted sensory input? Interest in this question is more than academic. The answer may help facilitate auditory learning by people who became deaf after learning language and later received a cochlear implant (a neural prosthesis that restores hearing through direct electrical stimulation of the auditory nerve). There is evidence that some cochlear implants (which provide information that is spectrally degraded to begin with) stimulate neurons with higher characteristic frequency than the acoustic frequency of the original stimulus. In other words, the stimulus is shifted in frequency with respect to what the listener expects to hear. This frequency misalignment may have a negative influence on speech perception by CI users. However, a perfect frequency-place alignment may result in the loss of important low frequency speech information. A trade-off may involve a gradual approach: start with correct frequency-place alignment to allow listeners to adapt to the spectrally degraded signal first, and then gradually increase the frequency shift to allow them to adapt to it over time. We used an acoustic model of a cochlear implant to measure adaptation to a frequency-shifted signal, using either the gradual approach or the “standard” approach (sudden imposition of the frequency shift). Listeners in both groups showed substantial auditory learning, as measured by increases in speech perception scores over the course of fifteen one-hour training sessions. However, the learning process was faster for listeners who were exposed to the gradual approach. These results suggest that gradual rather than sudden exposure may facilitate perceptual learning in the face of a spectrally degraded, frequency-shifted input.

---

© 2014 Elsevier B.V. All rights reserved.

<sup>1</sup>Corresponding author: Mario A. Svirsky, svirsky@nyu.edu, (646) 284 2457, Dept. of Otolaryngology-HNS, New York University School of Medicine, 550 First Avenue, NBV-5E5, New York, NY 10016.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

Cochlear implants are the first successful example of replacing a human sensory end organ with an electronic device. This accomplishment has been recently recognized with the awarding of the Lasker-DeBakey Clinical Medical Research Award to Blake Wilson, Graeme Clark, and Ingeborg Hochmair, three of the pioneers in the field. There are two main reasons why the cochlear implant is considered one of the major medical advances of the twentieth century. One of them is the impact it has had on the clinical treatment of hearing impairment and on the quality of life of hundreds of thousands of patients. These patients include adventitiously hearing impaired adults as well as children who were born profoundly deaf and whose sole auditory input was provided by the implant. The case of children is significant because in their case cochlear implants influence not only their ability to understand speech but also their ability to speak intelligibly and their development of oral language. Additionally, the widespread use of cochlear implants has caused a paradigm shift in a number of related fields by opening a number of new avenues for scientific and clinical pursuit. For example, it seems likely that the development of retinal prostheses has significantly benefitted in many different ways from the prior development of cochlear implants.

There are also numerous examples of important scientific questions that could not have been answered were it not for the availability of a large base of cochlear implant users. The study of sensitive periods for speech and language development is one of these examples. The sensitive period hypothesis states that the innate human ability to develop certain skills including language development and intelligible speech production decreases without early exposure to oral language. This is a hypothesis that cannot be easily tested with animal models because animals do not have speech and language. The definitive way to test the hypothesis would be to withdraw auditory input from children at birth and then measure their ability to develop speech language when hearing is restored after different periods of deprivation. This would be obviously unethical. However, cochlear implantation of children who are congenitally and profoundly deaf provides a unique opportunity to conduct an indirect test of the sensitive period hypothesis, and even makes it possible to estimate the length of such sensitive periods for different specific skills. The way this is done is by testing children who have received cochlear implants at different ages and therefore have been deprived of auditory input for different amounts of time (Tomblin et al., 2005, Geers et al, 2011, Tobey et al., 2013, Niparko et al., 2010, Svirsky et al., 2000, Svirsky et al., 2004b, Svirsky et al., 2007). Admittedly, this is not a perfect experiment because cochlear implants do not restore perfect hearing, but it is clear that the existence of cochlear implant technology made it possible to obtain a novel source of data to study the nature and time course of sensitive periods in language development. A second area of scientific study where cochlear implants have opened new doors is the study of the influence of auditory input on speech production. Cochlear implants make it possible to easily manipulate auditory input and measure the subtle effects on the listener's speech production. They also make it possible to completely deprive a listener of sound for periods ranging from minutes to several hours, to measure the changes that occur in speech production parameters in response to the deprivation, and immediately after auditory input is restored (Matthies et al.,

1994, Lane et al., 1997, Svirsky and Tobey, 1991, Svirsky et al., 1992). Cochlear implants make it possible to study changes in speech production after implantation in listeners who have been sound deprived for years or even decades (Perkell et al., 1992) and the resulting data can provide insights about underlying mechanisms in normal speech production (Lane et al., 1995, Perkell et al., 1995). Finally, a third area worth mentioning is that cochlear implants have had a profound influence on the study of speech perception using a very impoverished auditory input. The realization that a relatively primitive and distorted signal can be so useful in a pragmatic sense has given renewed impulse to studies that investigate the human brain's ability to understand speech in very adverse situations. Here we present a modest example of a study that does not involve cochlear implants directly but was clearly inspired by thinking about the mechanisms that postlingually deaf cochlear implant users might employ to understand speech.

It is well known that the human brain's ability to recognize patterns in spite of an extremely distorted input is very robust, at least when those distortions occur along a single dimension. In particular, humans can recognize speech in very adverse acoustic conditions that are well beyond the capabilities of the most sophisticated automatic speech recognition systems. Some well known examples include sinewave speech (where each formant frequency is represented by an amplitude and frequency modulated sinewave (Remez et al., 1981)), infinite peak clipping (where speech is converted to a signal that can assume only two values, one positive and one negative, and only the zero crossings of the original signal are preserved (Licklider et al., 1948)), and a four channel noise vocoder (where the input signal is filtered by four adjacent frequency channels and the output of each channel modulates a different noise band (Shannon et al., 1995)). However, there are indications that simultaneous distortion along two dimensions may be a more significant problem. See, for example, Fu and Shannon (1999), Rosen et al. (1999), Baskent and Shannon (2003), and Svirsky et al. (2012). The latter includes audio examples of a clear speech signal along with versions of the same signal that were degraded along one or two different acoustic dimensions.

Because auditory sensory aids impose different kinds of distortions, it is important to understand how listeners perceive degraded speech. This understanding may inform the design of next generations of such aids. For example, cochlear implants impose at least two kinds of signal distortion. One of them is the inevitable spectral degradation that is dictated by a relatively small number of independent stimulation channels. In postlingually deaf adults, cochlear implants may also impose a mismatch between the acoustic frequency of the input signal and the characteristic frequency of the neurons that are stimulated in response to that input. Why does this happen? There is mounting evidence that some models of cochlear implant electrode arrays are not inserted deeply enough to stimulate neurons whose characteristic frequency is under 300 Hz (Harnsberger et al. (2001), Carlyon et al. (2010), Svirsky et al. (2001), Svirsky et al. (2004a), Zeng et al. (2014)). One particularly compelling data set can be found in McDermott et al. (2009), who tested five users of the Nucleus cochlear implant (model CI24RE(CA)) before they had any experience with the device. These five subjects had some usable acoustic hearing that allowed them to match the percepts elicited by stimulating the most apical electrode to the percepts elicited by pure tones presented to the unimplanted ear. The tones that provided the best match ranged from

about 600 Hz to 900 Hz, suggesting that this device imposes an initial basalward frequency shift of one to two octaves (because the most apical electrode is stimulated in response to ambient sounds around 250 Hz but it actually “sounds like” a tone of 600 to 900 Hz, at least upon initial stimulation). Basalward shift imposes a conflict between the percepts caused by speech sounds and the internal representation of the same speech sounds that was learned by the same (postlingually deaf) listener when he had normal hearing. In other words, this is a discrepancy between novel sensory input and representations of speech sounds stored in long term memory.

Given the possibility of such basalward shift, designers of auditory prostheses are faced with two main options. The first one is to adjust the analysis filters of the speech processor to the estimated characteristic frequency of the stimulated neurons. This option has the advantage that no adaptation is required on the part of the listener but it comes at the expense of losing some low-frequency information that is important for speech perception. Another option is to accept the frequency misalignment and let listeners adapt to it over time. This is what is normally done with users of cochlear implants. This approach has the advantage that the most relevant speech information is delivered to the listener, but it requires adaptation to a frequency shift that can be quite severe in some individuals, and this may lead to incomplete adaptation in some of them. There are a number of studies showing that postlingually deafened cochlear implant users have a remarkable ability to adapt to the standard clinical frequency allocation tables, or even frequency tables that are further shifted from the standard (e.g., Skinner et al. (1995), Svirsky et al. (2001), McKay and Henshall (2002), Svirsky et al., (2004a), Gani et al. (2007)). Other studies have shown that the electric pitch sensation associated with activation of any given electrode adapts with experience, generally becoming more consistent with the frequency information provided by the cochlear implant processor at that electrode’s location (Svirsky et al., 2004; Reiss et al., 2007, 2014). Taken together, these studies support the feasibility of the second option, which is the standard approach in cochlear implant users. On the other hand, there are numerous studies suggesting that adaptation to standard frequency allocation tables may take a long time or may even be incomplete despite months or years of experience (Fu et al. (2002), Svirsky et al. (2004a), Francart et al. (2008), Sagi et al. (2010), Svirsky et al. (2012)). These studies tend to support the use of the first option listed above, either to facilitate and speed up the adaptation process or to enhance speech perception.

However, there is a third option: imposing the necessary frequency shift in small steps rather than all at once. This type of approach is reminiscent of the work on barn owls done by Knudsen and colleagues (1985a, 1985b). Barn owls rely on audiovisual information to home in on prey. Knudsen et al. found that baby barn owls fit with spectacles that shift the visual field by 23 degrees (and therefore impose a mismatch between auditory and visual input) acquire new auditory maps that match the visual input, after a few months. On the other hand, adult barn owls are largely unable to adapt in this manner. However, it was found that when the prismatic shift was imposed in small steps taken over time, adult barn owls showed much greater ability to adapt (Linkenhoker and Knudsen (2002), page 294. The authors commented that their results “show that there is substantially greater capacity for plasticity in adults than was previously recognized, and highlight a principled strategy for tapping this capacity”.

This is precisely what was attempted in the present study. Speech was degraded in two ways at the same time: by using a real time noise vocoder, and by using different amounts of frequency mismatch between the input analysis filters and the output noise bands of the vocoder. Pairs of subjects were matched for their ability to recognize vocoded speech and then assigned to one of two groups: a “standard” group, where the same frequency shift was kept constant through the 15-session experiment, and a “gradual” group, where the frequency shift was introduced in small steps over the first ten sessions. The goal was to test whether gradual introduction of a frequency shift helps or hinders perception of vocoded, frequency-shifted speech. We felt that our results would be potentially relevant for optimizing training with auditory sensory aids that impose a frequency shift to which postlingually deaf listeners must adapt, as is the case of cochlear implants and frequency-transposition hearing aids.

## Methods

### Subject selection

Twenty volunteers were screened to make sure they had normal hearing, defined as a pure tone average of less than 20 dB HL in each ear. All of them passed the screen and were tested in two separate pre-treatment sessions using closed-set consonant and vowel identification tests, as well as speechtracking. During the two pre-treatment sessions all speech stimuli were processed using an 8-channel noise vocoder where the analysis filters and synthesis filters were exactly matched (zero frequency shift). Normal hearing listeners can have a wide range of ability to understand degraded speech, so the goal was to obtain matched pairs of listeners with approximately similar abilities. There was indeed a wide range of abilities in our preliminary testing, with vowel identification scores averaged over the two pre-treatment sessions ranging from 17% to 93% correct across participants, consonant scores between 30% and 81% correct, and speechtracking scores between 55 and 85 words per minute. Seven matched pairs of participants (one male, thirteen females) were successfully obtained and each member of each pair was randomized to the “standard” group or the “gradual” group. The average difference in scores between the two groups was less than 1% for vowel identification, 2% for consonant identification, and 3 words per minute during speech tracking. The average of the absolute value of score differences between the groups was 3% for vowels, 5% for consonants, and 7 words per minute. This represents a very good match in the ability of each group to understand degraded speech prior to training. The average age of both groups was 23 years, with a range of 19 to 26 in the gradual group and 18 to 32 in the standard group.

### Signal Processing

Speech was processed in real time using a PC-based noise vocoder based on that described by Kaiser and Svirsky (2000). Audio was sampled with an Audigy sound card and ASIO drivers at 48,000 samples per second, downsampled to 24,000 samples per second, and anti-aliased using a low pass filter with a cutoff frequency of 11,000 Hz. The signal was then passed through a bank of eight analysis filters (fourth order elliptical IIR, implemented using the cascade form). Envelope detection was done at the output of each analysis filter by half wave rectification and second order low-pass filtering at 160 Hz. Each one of the eight

envelopes was used to modulate a noise band, which was obtained by passing white noise through a synthesis filter (fourth order elliptical IIR, just like the analysis filters). The total input-output latency of the system was less than 20 msec, which is a barely noticeable delay. This is important because it allows the use of the audiovisual speechtracking task (described below), where the listener must integrate auditory information with lipreading cues from a talker's face.

## Filters

Tables 1 and 2 show the frequency boundaries for the analysis and synthesis filters, for each one of the subject groups (standard and gradual). The synthesis filters were identical for both groups and were constant throughout the experiment, covering a frequency range of 854–11,000 Hz. This frequency range corresponds to intracochlear locations between 4 and 22 mm from the base of the cochlea (according to the Greenwood (1990) formula and assuming a cochlear length of 35 mm). The frequency range for the synthesis filters mentioned above was chosen because it approximates the characteristic frequency of the neurons that are closest to electrodes used by some cochlear implant systems, particularly Advanced Bionics and Cochlear (the standard Med-El electrode is longer, is typically inserted more deeply into the cochlea and therefore would correspond to lower frequencies than those used in this study). Analysis filters for the standard group were also constant throughout the experiment and they were modelled after the filters normally used in cochlear implant systems, covering a frequency range of 250–6800 Hz (in fact, the filter bandwidths were identical to those used in previous generations of Advanced Bionics 8-channel speech processors). Thus, subjects in the standard group were presented with an acoustic signal that was degraded in two ways: it was spectrally degraded by using an eight channel noise vocoder, and was frequency shifted by imposing a difference between the frequency range of the analysis filters and that of the synthesis filters. The standard group condition is intended as a model of standard clinical practice with auditory aids that may impose a frequency shift (i.e., accepting whatever mismatch may be imposed in the hopes that patients will adapt to it). Potential drawbacks are that the adaptation process may take time, and that the adaptation itself may not be complete.

In contrast, subjects in the gradual group initially used a set of analysis filters (shown in line 1 of Table 1) that were an exact match to the noise bands. Because the initial set of analysis filters covered the 854–11,000 Hz range there was no frequency mismatch, but this was done at the expense of eliminating all information below 854 Hz, which is very important for speech perception. Then, frequency mismatch and (in consequence) speech information between 250 and 854 Hz were introduced gradually, over several one-hour training sessions. As Table 1 shows, it is only at session 10 that both groups started using the same exact speech processing parameters. The “shift” column in Table 1 indicates the amount of mismatch between analysis and synthesis filters, measured in mm of cochlear displacement. The Greenwood formula was used to convert frequency differences in Hz to amounts of displacement along an average sized cochlea. As can be observed in Table 1, frequency shift was introduced at an initial rate of 1 mm per one-hour session for the first four sessions, and 0.5 mm per session thereafter until the maximum mismatch of 6.5 mm was reached. Figure 1 illustrates how the vowel /i/ (whose spectrum is shown in the top left panel) is affected by

different types of noise vocoders. The noise vocoder type most often used in the literature does an adequate job at preserving the rough spectral content of the original signal (top right panel). The bottom panels show the difference between the two testing groups in the study. Spectral information is roughly preserved for the standard group, but delivered to much higher frequency locations (see how this spectrum is similar to those in the top panels, but shifted to the right). In contrast, the initial spectral information for the gradual group is delivered to the right place but all spectral content up to 851 Hz has been eliminated, as can be appreciated by the absence of the first formant.

The present study did not attempt to model spectral warping of frequency-to-place maps, a third type of distortion that postlingually deafened adults with cochlear implants may experience due to uneven patterns of neural survival and other factors (see, for example, Ba Kent and Shannon (2003, 2004, 2006), Faulkner (2006)).

### Training/testing sessions

Both groups underwent fifteen one-hour sessions over the course of approximately one month (an average of twenty nine days with a standard deviation of seven days). The small differences across subjects in total time course for the experiment are very unlikely to have affected the results, because it has been shown that the type of perceptual learning examined in this study depends much more strongly on the amount of training than on the total training period, even when the latter changes by a factor of as much as five to one (Nogaki et al., 2007). Two ten minute blocks of audiovisual speechtracking (De Filippo and Scott, 1978) as well as audio-alone consonant and vowel recognition tests were done in every session. The material for the speechtracking task was the children's book "Danny, the Champion of the World" (Dahl, 1978). Two lists of the Central Institute for the Deaf (CID) open set sentence test (Silverman and Hirsch, 1955) were administered per session but only in sessions 1, 5, 9, 13 and 15, because there were not enough sentence lists to run that test in every session without repeats. All tests were conducted at 70 dB SPL, C-scale.

Speechtracking is a task where the experimenter reads a passage, phrase by phrase, and the listener has to repeat verbatim what he heard. The phrase is repeated or broken down into shorter segments until the listener understands what was said. This was done in the audiovisual condition by having the experimenter and listener sit on opposite sides of a sound booth window. All sessions within the one month training period were conducted by the same experimenter, who was a female talker. The experimenter's speech was processed in real time with a delay of less than 20 ms, and the audio signal was routed into the booth where listeners heard it using headphones while seeing the experimenter's face outside the booth. This task was selected in part due to its ecological validity because cochlear implant users generally have access to both visual and auditory information in their everyday interactions. Also, the audiovisual presentation helped avoid floor effects that might have initially occurred in the audio-alone condition. Thus, the speechtracking task provided feedback that could be useful for the listener's learning process.

The consonant and vowel tests (Tyler et al., 1987) were closed-set and conducted in the audio-alone condition. Stimuli included three productions each of sixteen consonants in "a-consonant-a" context and five productions each of nine vowels in "h-vowel-d" context by a

female talker. After each consonant or vowel response, the listener was told whether the response was correct or incorrect. If the response was incorrect, the correct response was visually displayed on the computer screen and sound playback was repeated. Again, this feedback was provided as part of the training process by which listeners learned to recognize the degraded speech signal. Just like with the speechtracking task, there were no separate testing and training sessions. All experimental sessions were done with feedback and thus provided an opportunity for training, and the resulting percent correct data was the outcome of this part of the study. The three training/testing tasks were chosen in part because different forms of phoneme training and sentence recognition using speechtracking have proven effective to help listeners improve recognition of spectrally shifted speech, and have been shown to generalize to speech perception tasks other than that which was explicitly trained (Fu et al., 2005; Rosen et al., 1999).

After the end of the one month training period, subjects underwent one additional session where they did the speechtracking task with another talker, different from the experimenter who conducted the sessions during the training period.

### Statistics

First, a linear regression was done for each one of the four outcome measures (vowels, consonants, sentences, speechtracking) as a function of testing interval. This was done separately for the standard and gradual groups. The way that different subjects within a matched pair were assigned to different groups represents a randomized block design for the Group factor (which can assume two levels, standard or gradual), and measures were repeated on the same subjects at every session (from session 1 to session 15). Randomized block data are analyzed the same way as repeated measures data, so each outcome measure was analyzed separately with a 2-way repeated measures ANOVA, after each data set passed a Wilk-Shapiro normality test and an equal variance test. Post-hoc pairwise multiple comparisons were done using the Holm-Sidak method with an overall significance level of 0.05. A similar analysis was conducted to assess generalization of speechtracking to a different talker. In this case only three testing sessions were used: the first and final session with the first talker (sessions 1 and 15) and the session with the new talker (session 16).

### Results

Figure 2 shows the average speechtracking data for both groups, with white circles representing the gradual group and black circles for the standard group. The statistical analysis revealed a significant group by interval interaction ( $p < 0.001$ ), consistent with the trends that can be observed in the figure. The standard group suffered a large initial drop with respect to their performance with an unshifted noise vocoder and it took about nine testing sessions for them to catch up to the gradual group. On the other hand, the gradual group reached near asymptotic performance during the first session. Post-hoc tests showed a significant advantage for the gradual group in sessions 1, 2, 4, 5, 6, and 7. The consonant scores are shown in Figure 3 and they largely parallel the speech tracking results. There was a significant group by interval interaction ( $p < 0.001$ ). The standard group suffered a much more significant drop from the PRE session to session 1 than the gradual group, resulting in



a sizeable initial advantage for the latter group (51% correct vs. 27% correct). However, the standard group caught up with the gradual group by the seventh test interval. Between-group differences were statistically significant only for the first three test intervals. Sentence identification scores, shown in Figure 4, were also consistent with the speechtracking and consonant identification scores: there was a significant group by interval interaction ( $p < 0.01$ ) and an advantage for the gradual group at test interval 1 ( $p < 0.05$ ) that disappeared at later test intervals. Recall that this measure was only used in a subset of the test intervals to avoid having to use a list of sentences more than once. Lastly, vowel scores are shown in Figure 5. Unlike with the other outcome measures, there were no statistically significant differences in percent correct scores between the two groups at any of the test intervals.

Linear regressions for the standard group showed significant increases as a function of testing interval for each outcome measure ( $p < 0.001$  for vowels, consonants and speechtracking, and  $p = 0.015$  for CID sentences). In contrast, the gradual group showed significant increases as a function of testing interval only for the vowel scores ( $p < 0.001$ ). Taken together, these results suggest that the gradual group was able to achieve asymptotic speech perception scores much earlier than the standard group for the consonant, sentence, and speechtracking tasks but not for the vowel task, where there was no difference between the two groups. In any case there were clearly no differences in scores between the two groups after session 7. The final levels were the same for both groups and for all four speech measures.

However, the gradual group had an advantage when performing speechtracking with a novel talker. Figure 6 shows the average speechtracking scores for each group, and the asterisks indicate statistically significant differences between groups ( $p < 0.05$ ). The first two pairs of bars simply repeat what was shown in Figure 2 for testing sessions 1 and 15: a significant initial advantage for the gradual group that disappears by the fifteenth testing session. The bars to the right show that the gradual group still outperformed the standard group when speechtracking was done with a new talker in session 16. There were no significant differences across the three test intervals within the gradual group, consistent with the observation that these subjects were able to do the task at the same performance level with the novel talker as with the original talker. In contrast, speechtracking scores for subjects in the standard group were significantly different for each one of the three sessions. Scores obtained with the second talker were significantly higher than those obtained with the first talker in the first session but significantly lower than those obtained with the first talker in the fifteenth session. This is consistent with the possibility that the process of generalizing speech perception skills to new talkers was more robust for the gradual group than for the standard group.

## Discussion

A first important (if unsurprising) observation is that speech perception scores tended to increase with training. However, the main question that the present study tried to answer was whether there were any differences between the gradual and the standard approach to help human listeners adapt to a spectrally degraded and frequency shifted signal. There are different reasons why one might expect that either approach might be better. On one hand,

the standard approach has the advantage that listeners train right from the start using a single consistent speech processing strategy. On the other hand, the gradual approach may facilitate the listener's job initially by leveraging his knowledge about the speech signal and his expectations about how speech should sound, adjusting those expectations gradually rather than using a "sink or swim" approach. Results showed that the null hypothesis could be rejected. In other words, results were not identical for the two groups. There was an initial advantage for the gradual group for some of the tests (although not for vowels), and this advantage disappeared after seven sessions at most. Another difference across groups is that even after listeners in the standard group caught up with those in the gradual group, the latter retained an advantage in speechtracking with a novel talker. Taken together, these results suggest that gradual exposure to frequency mismatch may result in faster improvement when trying to perceive a degraded speech signal (although this advantage may be fleeting) and perhaps in other advantages after asymptotic speech perception levels are reached. The present results are consistent with those from Li et al. (2009), who found that listeners exposed exclusively to a severely frequency shifted signal showed much less improvement in speech perception scores than listeners who were also exposed to some amount of moderately shifted speech during the course of their experiment.

It is interesting that the vowel scores (unlike the consonant, speechtracking, and sentence scores) showed no differences across the two groups. This may be because in the case of vowels the information that is initially lost when using a gradual approach is a very significant part of the overall phonetic information. Examination of the first line of Table 1 shows that almost all the first formant frequency range (up to 854 Hz) is initially unavailable to the gradual group. Perhaps the tradeoff of reduced information in exchange for less frequency shift is favourable in the case of consonants because there are many other acoustic cues in addition to first formant frequency, whereas the tradeoff is roughly even in the case of vowels.

This study was at least partly inspired by the kind of speech processing that is used in cochlear implants, and the standard approach represents an approximation of current clinical practice. When fitting the speech processor of an adult cochlear implant patient there are usually no attempts to measure the extent of a possible frequency mismatch, in large part because the standard fitting software provided by cochlear implant manufacturers does not include tools for that kind of measurement. The present results suggest there may be some value in trying to measure frequency mismatch in adult cochlear implant users and there are experimental approaches that might help in this regard (Jethanamest et al., 2010, Wakefield et al., 2005, Holmes et al., 2012). On the other hand, and with the exception of the measures of generalization to a new talker, the differences found in this noise vocoder study were not long lasting. This means that the gradual approach may or may not be all that advantageous to real cochlear implant users, in a pragmatic sense, when we consider the additional work that it requires. The only way to settle this question would be with a follow up study in the real clinical population.

There are two important caveats. First, the present results may have implications for the treatment for postlingually deaf adults who have to adapt to a frequency shift but it does not have any implications whatsoever for the treatment of congenitally deaf children with

cochlear implants. By definition these children cannot suffer from a frequency mismatch between the stimulation they receive from the implant and their expectations about how speech should sound, because in their case such expectations are developed based on input provided exclusively by the implant. Second, conclusions should be tentative even in the case of postlingually deaf adults because this study was conducted using two kinds of models. The one-month training regimen used in the study may be thought of as a model of the adaptation that postlingually deaf cochlear implant users undergo after initial stimulation, and perception of speech by normal listeners hearing through a noise vocoder has long been considered a model of speech perception by cochlear implant users. However, new studies in single sided deafness patients comparing the percepts provided by cochlear implants to those provided by acoustic models show that some vocoders sound very similar to the cochlear implant and others do not. The situation is further complicated by the fact that the most accurate acoustic models are different across individual cochlear implant users (Svirsky et al., 2013).

This study attempted to answer some questions about the way in which adult listeners may adapt to a distorted speech input. The study did not involve any actual cochlear implant users or any clinical speech processors, but was clearly inspired by the experience of postlingually deaf cochlear implant users. Thus, it represents a modest but concrete example of the way cochlear implants have caused not just a clinical but a scientific paradigm shift, by prompting scientists to pursue basic questions that may have translational implications in this new population of auditory neural prosthesis users.

## Acknowledgments

This study was supported by funding from NIH grant DC03937 (P.I.: Svirsky).

## References

- Ba kent D, Shannon RV. Speech recognition under conditions of frequency-place compression and expansion. *Journal of the Acoustical Society of America*. 2003; 113:2064–2076. [PubMed: 12703717]
- Ba kent D, Shannon RV. Frequency-place compression and expansion in cochlear implant listeners. *The Journal of the Acoustical Society of America*. 2004; 116(5):3130–3140. [PubMed: 15603158]
- Ba kent D, Shannon RV. Frequency transposition around dead regions simulated with a noiseband vocoder. *The journal of the Acoustical Society of America*. 2006; 119(2):1156–1163. [PubMed: 16521776]
- Carlyon RP, Macherey O, Frijns JH, Axon PR, Kalkman RK, et al. Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *J Assoc Res Otolaryngol*. 2010; 11:625–640. [PubMed: 20526727]
- Dahl, R. *Danny, the Champion of the World*. Publisher: Bantam Books; New York City: 1978.
- De Filippo CL, Scott BL. A method for training the reception of ongoing speech. *Journal of the Acoustical Society of America*. 1978; 63:1186–1192. [PubMed: 649877]
- Faulkner A. Adaptation to distorted frequency-to-place maps: Implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation. *Audiology and Neurotology*. 2006; 11(Suppl 1):21–26. [PubMed: 17063007]
- Francart T, Brokx J, Wouters J. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. *Audiol Neurootol*. 2008; 13:309–319. [PubMed: 18391567]

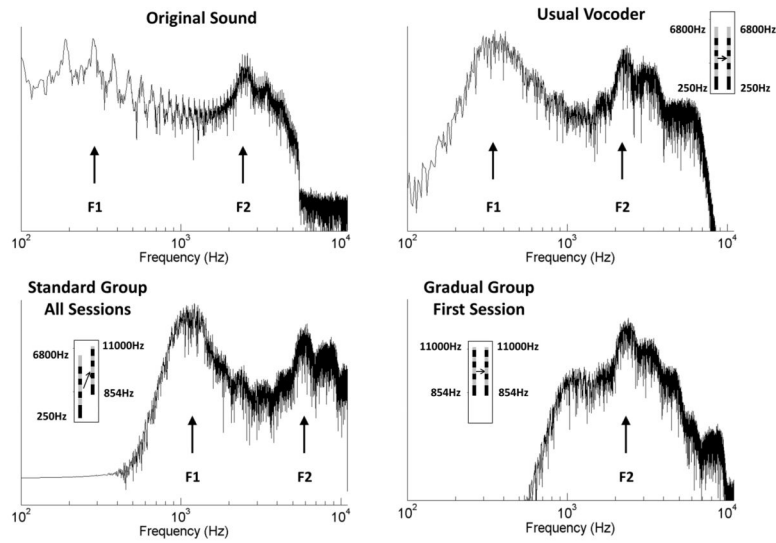
- Fu Q-J, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America*. 1999; 105:1889–1900. [PubMed: 10089611]
- Fu QJ, Shannon RV, Galvin JJ III. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *The Journal of the Acoustical Society of America*. 2002; 112(4):1664–1674. [PubMed: 12398471]
- Fu QJ, Nogaki G, Galvin JJ 3rd . Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *J Assoc Res Otolaryngol*. 2005; 6(2):180–9. [PubMed: 15952053]
- Gani M, Valentini G, Sigrist A, Kós MI, Boëx C. Implications of deep electrode insertion on cochlear implant fitting. *Journal of the Association for Research in Otolaryngology*. 2007; 8(1):69–83. [PubMed: 17216585]
- Geers AE, Strube MJ, Tobey EA, Pisoni DB, Moog JS. Epilogue: factors contributing to long-term outcomes of cochlear implantation in early childhood. *Ear & Hearing*. 2011; 32(1 Suppl):84S–92S. [PubMed: 21832892]
- Greenwood DD. A cochlear frequency-position function for several species--29 years later. *J Acoust Soc Am*. 1990; 87(6):2592–605. [PubMed: 2373794]
- Harnsberger JD, Svirsky MA, et al. Perceptual “vowel spaces” of cochlear implant users: implications for the study of auditory adaptation to spectral shift. *Journal of the Acoustical Society of America*. 2001; 109(5 Pt 1):2135–45. [PubMed: 11386565]
- Holmes AE, Shrivastav R, Krause L, Siburt HW, Schwartz E. Speech based optimization of cochlear implants. *Int J Audiol*. 2012 Nov; 51(11):806–16. [PubMed: 22978753]
- Jethanamest D, Tan CT, Fitzgerald MB, Svirsky MA. A new software tool to optimize frequency table selection for cochlear implants. *Otol Neurotol*. 2010; 31(8):1242–7. [PubMed: 20729774]
- Kaiser, AR.; Svirsky, MA. Using a personal computer to perform real-time signal processing in cochlear implant research. *Proceedings of the IXth IEEE-DSP Workshop; Hunt, TX, USA.. 2000.*
- Knudsen EI, Knudsen PF. Vision guides the adjustment of auditory localization in young barn owls. *Science*. 1985a; 230(4725):545–8. [PubMed: 4048948]
- Knudsen EI. Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl. *J Neurosci*. 1985b; 5(11):3094–109. [PubMed: 4056865]
- Lane H, Wozniak J, Matthies ML, Svirsky MA, Perkell JS. Phonemic resetting versus postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time. *Journal of the Acoustical Society of America*. 1995; 98(6):3096–106. [PubMed: 8550935]
- Lane H, Wozniak J, Matthies ML, Svirsky MA, Perkell JS, O’Connell M, Manzella J. Changes in sound pressure and fundamental frequency contours following changes in hearing status. *Journal of Acoustical Society of America*. 1997; 101(4):2244–2252.
- Li T, Galvin JJ III, Fu QJ. Interactions Between Unsupervised Learning and the Degree of Spectral Mismatch on Short-Term Perceptual Adaptation to Spectrally Shifted Speech. *Ear and Hearing*. 2009; 30(2):238–249. [PubMed: 19194293]
- Licklider JCR, Irwin Pollack. Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech. *The Journal of the Acoustical Society of America*. 1948; 20:42–51.
- Linkenhoker BA, Knudsen EI. Incremental training increases the plasticity of the auditory space map in adult barn owls. *Nature*. 2002; 419(6904):293–6. [PubMed: 12239566]
- Matthies ML, Svirsky MA, Lane H, Perkell JS. A preliminary study of the effects of cochlear implants on the production of sibilants. *Journal of the Acoustical Society of America*. 1994; 96(3):1367–73. [PubMed: 7963001]
- McDermott H, Sucher C, Simpson A. Electro-acoustic stimulation acoustic and electric pitch comparisons. *Audiology and Neurotology*. 2009; 14(suppl 1):2–7. [PubMed: 19390169]
- McKay CM, Henshall KR. Frequency-to-electrode allocation and speech perception with cochlear implants. *The Journal of the Acoustical Society of America*. 2002; 111(2):1036–1044. [PubMed: 11863160]
- Niparko JK, Tobey EA, Thal DJ, Eisenberg LS, Wang NY, Quittner AL, Fink NE. CDaCI Investigative Team. *JAMA*. 2010; 303(15):1498–506. [PubMed: 20407059]

- Nogaki G, Fu QJ, Galvin JJ 3rd . Effect of training rate on recognition of spectrally shifted speech. *Ear Hear.* 2007; 28(2):132–40. [PubMed: 17496666]
- Perkell JS, Lane H, Svirsky MA, Webster J. Speech of cochlear implant patients: A longitudinal study of vowel production. *Journal of the Acoustical Society of America.* 1992; 91(5):2961–78. [PubMed: 1629489]
- Perkell JS, Matthies ML, Svirsky MA, Jordan MI. Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics.* 1995; 23:23–35.
- Reiss LA, Turner CW, Erenberg SR, et al. Changes in pitch with a cochlear implant over time. *J Assoc Res Otolaryngol.* 2007; 8:241–257. [PubMed: 17347777]
- Reiss LA, Turner CW, Karsten SA, et al. Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation. *Neuroscience.* 2014; 256:43–52. [PubMed: 24157931]
- Remez RE, Rubin PE, Pisoni DB, Carell TD. Speech perception without traditional speech cues. *Science.* 1981; 212:947–950. [PubMed: 7233191]
- Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America.* 1999; 106:3629–3636. [PubMed: 10615701]
- Sagi E, Fu QJ, Galvin JJ III, Svirsky MA. A model of incomplete adaptation to a severely shifted frequency-to-electrode mapping by cochlear implant users. *Journal of the Association for Research in Otolaryngology.* 2010; 11(1):69–78. [PubMed: 19774412]
- Shannon RV, Zeng F-G, Wygonski J, Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304.
- Silverman SR, Hirsh IJ. Problems related to the use of speech in clinical audiometry. *Ann Otol Rhinol Laryngol.* 1955; 64(4):1234–44. [PubMed: 13283495]
- Skinner MW, Holden LK, Holden TA. Effect of frequency boundary assignment on speech recognition with the speak speech-coding strategy. *The Annals of otology, rhinology & laryngology Supplement.* 1995; 166:307.
- Svirsky MA, Tobey EA. Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users. *Journal of the Acoustical Society of America.* 1991; 89(6):2895–2904. [PubMed: 1918629]
- Svirsky MA, Lane H, Perkell JS, Webster J. Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *Journal of the Acoustical Society of America.* 1992; 92(3):1284–1300. [PubMed: 1401516]
- Svirsky MA, Robbins AM, Kirk KI, Pisoni DB, Miyamoto RT. Language development in profoundly deaf children with cochlear implants. *Psychological Science.* 2000; 11 (2):153–158. [PubMed: 11273423]
- Svirsky MA, Silveira A, et al. Auditory learning and adaptation after cochlear implantation: a preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Otolaryngol.* 2001; 121(2):262–5. [PubMed: 11349792]
- Svirsky MA, Silveira A, et al. Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngol.* 2004a; 124(4):381–6. [PubMed: 15224858]
- Svirsky MA, Teoh SW, Neuburger H. Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. 2004b; 9(4):224–233.
- Svirsky MA, Chin SB, Jester A. The effects of age at implantation on speech intelligibility in pediatric cochlear implant users: Clinical outcomes and sensitive periods. *Audiological Medicine.* 2007; 5:293–306.
- Svirsky MA, Fitzgerald MB, Neuman A, Sagi E, Tan CT, Ketten D, Martin B. Current and planned cochlear implant research at New York University Laboratory for Translational Auditory Research. *J Am Acad Audiol.* 2012; 23(6):422–37. [PubMed: 22668763]
- Svirsky, MA.; Ding, N.; Sagi, E.; Tan, C-T.; Fitzgerald, MB.; Glassman, EK.; Seward, K.; Neuman, A. Validation of Acoustic Models of Auditory Neural Prostheses. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing;* 2013; 2013. p. 8629-8633.
- Tobey EA, Thal D, Niparko JK, Eisenberg LS, Quittner AL, Wang NY. CDaCI Investigative Team. *International Journal of Audiology.* 2013; 52(4):219–29. [PubMed: 23448124]

- Tomblin JB, Barker BA, Spencer LJ, Zhang X, Gantz BJ. The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers. *J Speech Lang Hear Res.* 2005; 48(4):853–67. [PubMed: 16378478]
- Tyler, RS.; Preece, JP.; Lowder, MW. Laser Videodisc and Laboratory Report. Department of Otolaryngology-Head and Neck Surgery, University of Iowa at Iowa City; 1987. The Iowa audiovisual speech perception laser videodisc.
- Wakefield GH, van den Honert C, Parkinson W, Lineaweaver S. Genetic algorithms for adaptive psychophysical procedures: recipient-directed design of speech-processor MAPs. *Ear Hear.* 2005; 26(4 Suppl):57S–72S. [PubMed: 16082268]
- Zeng F-G, Tang Q, Lu T. Abnormal Pitch Perception Produced by Cochlear Implant Stimulation. *PLoS ONE.* 2014; 9(2):e88662. [PubMed: 24551131]

### Highlights

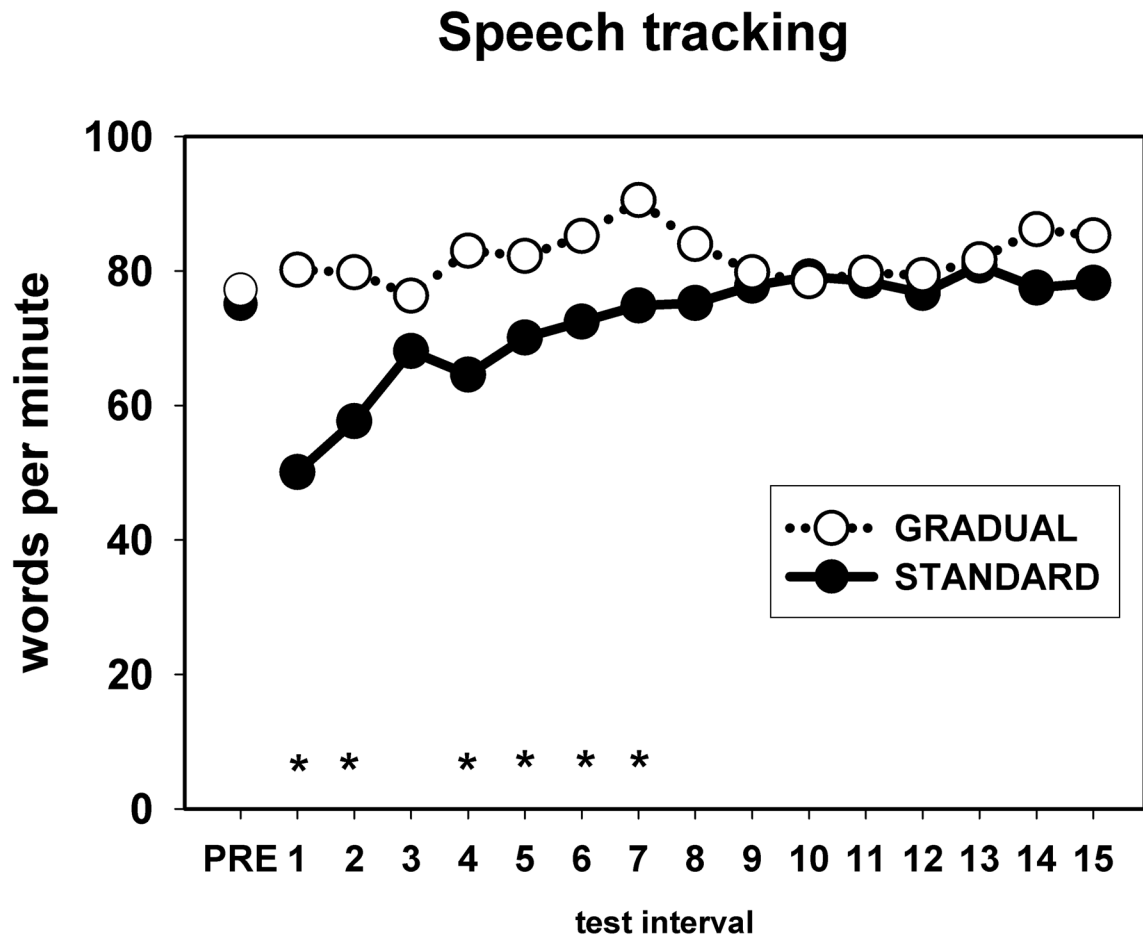
- We study how humans adapt to spectrally degraded, frequency shifted auditory input.
- Imposing the frequency shift in gradual small steps makes speech perception improve faster.
- The gradual approach also results in better generalization to understand the speech of novel talkers.



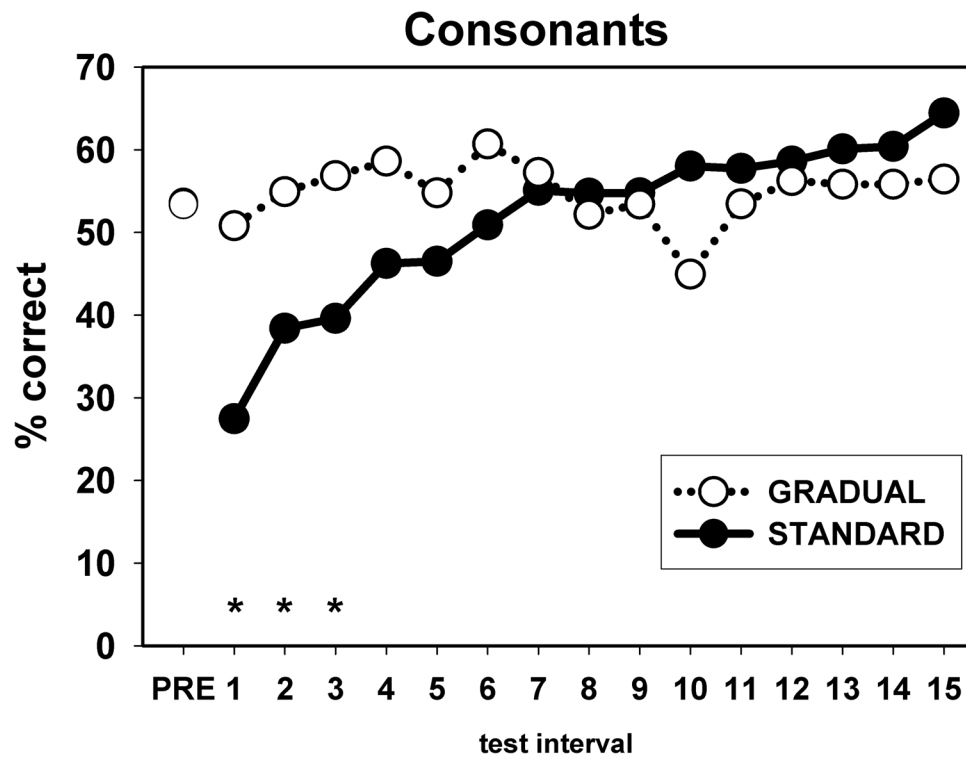
**Figure 1.**

The top left panel shows the spectrum of the vowel /i/ produced by a male talker and the three other panels show the output of various noise vocoders, with inset graphs that illustrate which specific vocoder configuration was used. The approximate location of the first formant frequency (F1) and second formant frequency (F2) is indicated in each panel, if the formant is present in the spectrum. The top right panel shows how spectral content is roughly preserved in the type of noise vocoder most commonly used in the cochlear implant literature (but not in this study). The values of F1 and F2 are largely unaffected. Both F1 and F2 are present at the output of the noise vocoder used for the standard group (bottom left) but delivered to a much higher frequency location. The noise vocoder used in the first session for the gradual group (bottom right) preserves F2 and sends it to the correct frequency location, but at the expense of eliminating much F1 information.

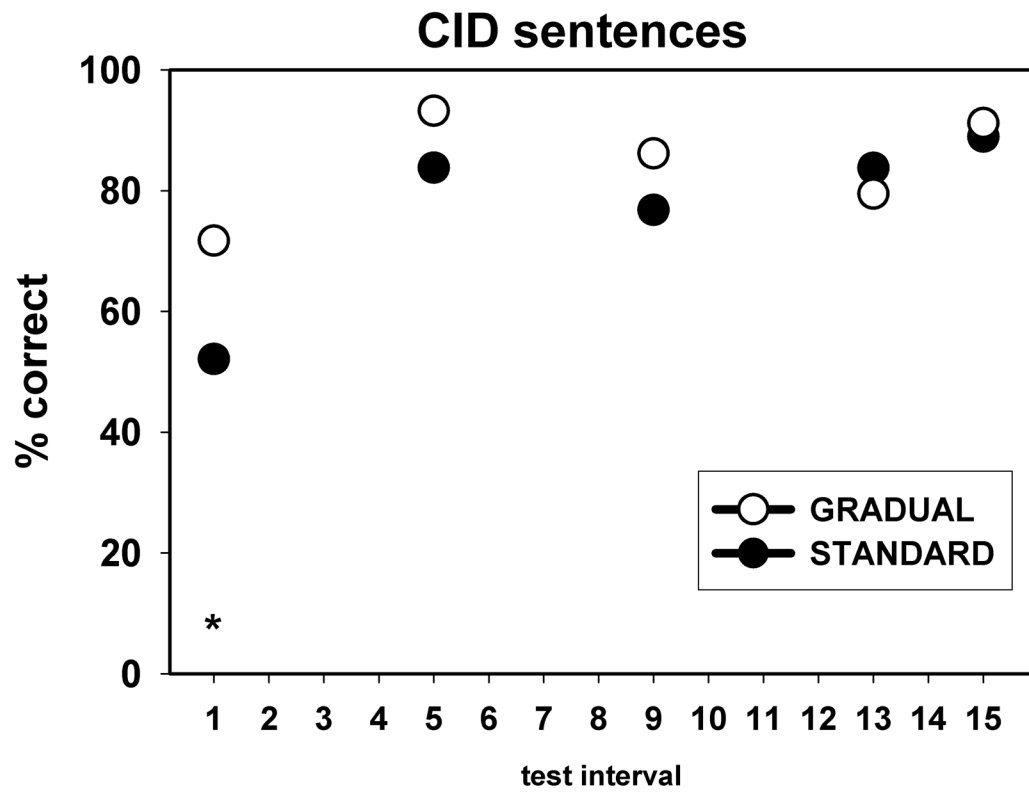




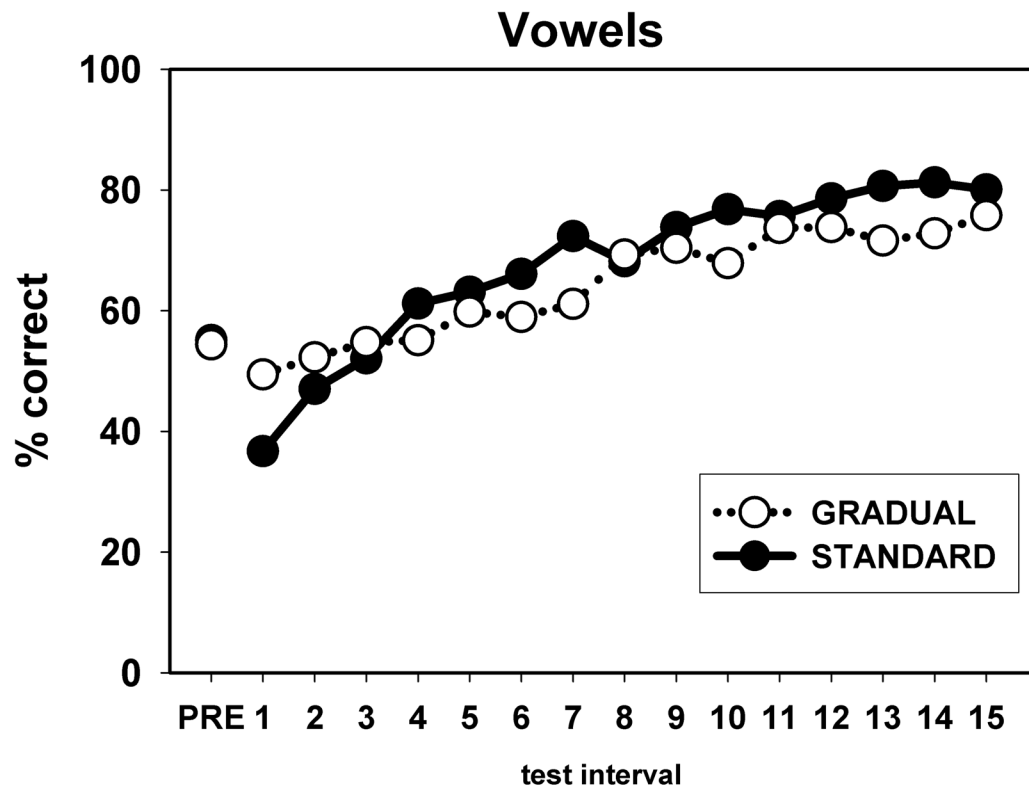
**Figure 2.** Average speechtracking scores for the gradual and experimental groups at each test interval. The PRE interval refers to the two speech testing sessions that were used to obtain pairs of subjects who were matched in their ability to understand speech with an unshifted noise vocoder. Asterisks indicate the test intervals where the gradual group's scores were significantly higher than the standard group's ( $p < 0.05$ ).



**Figure 3.** Average consonant scores for the gradual and experimental groups at each test interval. Just like with the speechtracking scores, the gradual group has an initial advantage and the standard group catches up after several testing sessions. Asterisks indicate significant between-group differences ( $p < 0.05$ ).

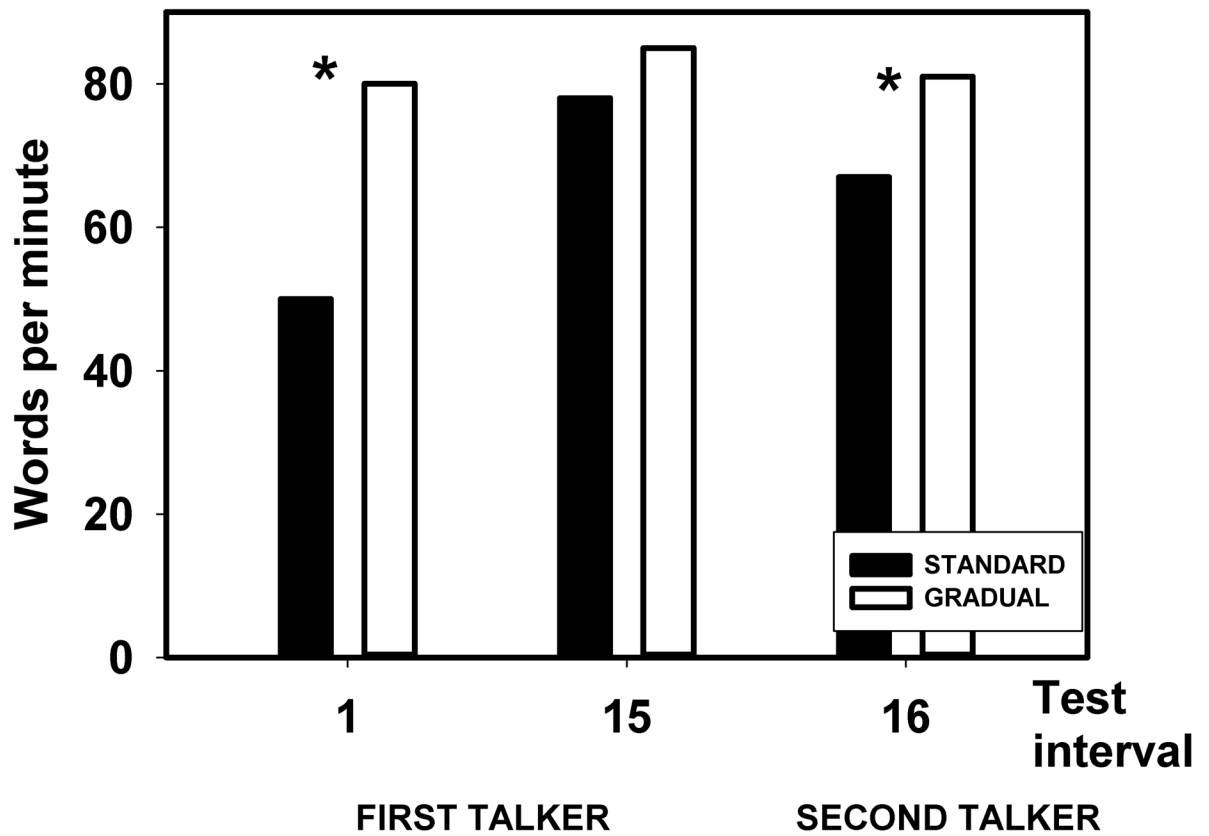


**Figure 4.** Average CID sentence scores for the gradual and experimental groups at each test interval. Just like with the speechtracking and consonant scores, the gradual group has an initial advantage at test interval 1 ( $p < 0.05$ ). There are no significant differences between groups at the later test intervals.



**Figure 5.** Average vowel scores for the gradual and experimental groups at each test interval. In contrast with the other outcome measures, there weren't any statistically significant differences between the two groups at any test interval.

## Speechtracking- Generalization



**Figure 6.** Speechtracking scores for the standard and gradual groups (black and white bars respectively). Asterisks indicate statistically significant differences between groups ( $p < 0.05$ ). The gradual group initially outperformed the standard group but the difference was no longer significant by the fifteenth session. The gradual group also had an advantage in session 16, which was conducted with a novel talker.

Analysis filter boundaries for the standard and gradual groups for each session. The “Shift” column indicates the amount of frequency mismatch between the analysis filters in this table and the synthesis filters in Table 2, expressed in mm of displacement along the cochlea.

**Table 1**

	Gradual Group Session numbers	Standard Group Session numbers	Shift (mm)	Analysis filter frequency boundaries (Hz)										
				854	1467	2032	2732	3800	5150	6622	9568	11000		
1	-	-	0	854	1467	2032	2732	3800	5150	6622	9568	11000		
2	-	-	1	722	1257	1749	2358	3288	4464	5746	8312	11000		
3	-	-	2	608	1073	1502	2032	2842	3866	4984	7218	11000		
4	-	-	3	508	914	1287	1749	2454	3346	4319	6266	11000		
5	-	-	4	421	774	1099	1502	2116	2893	3740	5435	9673		
6	-	-	4.5	382	711	1015	1390	1964	2689	3480	5062	9016		
7	-	-	5	345	653	936	1286	1822	2499	3237	4713	8404		
8	-	-	5.5	311	598	862	1190	1684	2320	3009	4387	7831		
9	-	-	6	279	547	794	1099	1565	2154	2798	4084	7298		
10-15	1-15	-	6.5	250	500	730	1015	1450	2000	2600	3800	6800		

**Table 2**

Synthesis filter boundaries for the gradual and constant shift groups

Gradual Group Session numbers	Standard Group Session numbers	Synthesis filter frequency boundaries (Hz)								
1-15	1-15	854	1467	2032	2732	3800	5150	6622	9568	11000