Summer 5-17-2012

Intelligibility and the Sinewave Speech Carrier Phrase Effect

Danielle R. Werle
Western Michigan University

THE CARL AND WINIFRED LEE HONORS COLLEGE

CERTIFICATE OF ORAL DEFENSE OF HONORS THESIS

Danielle Werle, having been admitted to the Carl and Winifred Lee Honors College in the fall of 2008, successfully completed the Lee Honors College Thesis on May 17, 2012.

The title of the thesis is:

Intelligibility and the Sinewave Speech Carrier Phrase Effect

Dr. James M. Hillenbrand, Speech Pathology and Audiology

Dr. Stephen Tasko, Speech Pathology and Audiology

Dr. Helen Sharp, Speech Pathology and Audiology

1903 W. Michigan Ave., Kalamazoo, MI 49008-5244
PHONE: (269) 387-3230 FAX: (269) 387-3903
www.wmich.edu/honors

Intelligibility and the Sinewave Speech Carrier Phrase Effect

Undergraduate Thesis

Dr. James Hillenbrand, Committee Chair

Danielle Werle

Western Michigan University

May 17, 2012

Abstract

Several studies have examined the perception of sinewave speech. Early work showed that listeners could transcribe sinewave sentences. Recent work on sinewave vowels found that listeners could recognize sinewave speech at the phonetic level with moderate (about 50-55%) accuracy. Sinewave vowel intelligibility scores improve significantly when preceded by a sinewave carrier phrase, an effect that disappears when the carrier phrase is removed. The present study was designed to examine this carrier phrase enhancement effect and analyze whether listeners require the carrier phrase to be intelligible to produce increased intelligibility scores. Sinusoidal carrier phrases and /hVd/ syllables were generated from recordings of an adult male and presented to listeners. Carrier phrases were recorded in both English and Japanese – a language unfamiliar to the 43 participants who served as listeners. All participants were students who had completed an introductory phonetics course. The conditions containing the Japanese carrier phrases resulted in intelligibility rates between 48 and 50%, roughly the same as the sinewave syllables being presented in isolation. The condition resulting in the highest intelligibility rate, 77%, included a single English carrier phrase, "The next word on the list is…" repeated on each of the 240 trials. The results indicate that the carrier phrase enhancement effect occurs only in the presence of an intelligible carrier phrase to signal to the listener to process the utterance as speech.

Introduction

Speech perception has been the subject of a great deal of research over the years. In natural speech, listeners are presented with a multitude of visual and acoustic information from the speaker which they use to identify and comprehend utterances. To assess which auditory components of a signal listeners are attending to when they process speech, researchers have used sinewave synthesized speech – signals made up of only three sinusoids. Studies examining the perception of sinewave speech, beginning with Remez, Rubin, Pisoni and Carrell (1981) have shown that, despite numerous acoustic differences between sinewave speech and natural speech, listeners are still able to transcribe sinewave sentences with moderate accuracy. The present study further investigates factors which control the intelligibility of sinewave speech.

Sinewave speech signals used in all of these studies differ from natural speech in several ways. The sinewave synthesized speech is merely a mix of three sinusoids of the same frequencies and amplitudes as the center frequencies of the lowest three formants in a speech sample (Figure 1). Since the component sinusoids are not harmonically related to one another, sinewave speech signals are aperiodic. Natural speech, on the other hand, is full of complex periodic and aperiodic elements due to vocal fold vibration as well as noise generated by unvoiced speech sounds. Additionally, some speech sounds, such as voiced fricatives and affricates, contain *both* periodic and aperiodic elements. Sinewave speech contains very narrow formant peaks compared to the broadband formants found in natural speech. The resulting sound is very unusual and dramatically different in quality than natural speech.

Remez et al. (1981) presented the sinewave synthesized sentence, “Where were you a year ago?” spoken by an adult male to 18 listeners. Three conditions were used in this study, but of most relevance to this paper are instructional conditions A and B. In condition A, listeners were not told anything about the utterance, and were simply asked to report their spontaneous impressions of what they were hearing. Remarkably, two accurately transcribed the sentence and about half chose speech-like answers such as human vocalizations and human or artificial speech. In condition B, a new set of 18 listeners were informed that the signal they were about to hear was speech, and were asked to transcribe the utterance to the best of their ability. Ten listeners reported hearing no sentence at all, even with the prompt. Overall, however, intelligibility rates were noticeably higher. Nine individuals were able to transcribe the sentence completely and several could transcribe parts or most of the utterance. It is important to note that the change from condition A to condition B consisted solely of the informing of the listeners that they were about to hear speech. The authors concluded that with the addition of this instruction, listeners were able to phonetically perceive the signals they were listening to. They asserted that listeners were able to direct their attention to the phonetic properties of the sinusoidal signal, even in the absence of traditional speech cues.

While sinewave sentences are clearly intelligible, the use of meaningful, well-formed sentences makes it impossible to determine the degree to which listeners are relying on phonetic information or utilizing a higher level of linguistic knowledge. Though what Remez et al. (1981) refer to as the ‘traditional speech cues’ are stripped away in sinewave sentences, an argument cannot be made for the amount of *phonetic* information being conveyed without measuring the intelligibility of sinewave signals at a *phonetic* level. In a subsequent study focused on testing vocal tract normalization, Remez, Rubin, Nygaard, and Howell (1987) measured vowel intelligibility of four vowels in /bVd/ format. These utterances were preceded by a carrier phrase, “Please say what this word is.” In their control condition, the carrier phrase was generated from the same speaker who recorded the vowel choices. Results of this condition showed that listeners could identify the syllables with 60% accuracy. While this figure is quite low, it shows that listeners can, in fact, obtain phonetic information from sinewave speech.

A more comprehensive study of the perception of sinewave vowels was later performed by Hillenbrand, Clark, and Baer (2011). By assessing vowel instead of sentence intelligibility, the consideration about the use of higher linguistic knowledge may be eliminated as sounds are presented in isolation and listeners must rely entirely on perceptual pattern-matching mechanisms operating exclusively at the phonetic level. In noting that participants in Remez et al.’s (1981) study produced more accurate results simply by being told to listen for speech, Hillenbrand, Clark, and Baer (2011) focused on three types of training procedures intended to increase participants’ accuracy in perceiving the sinewave vowel’s relationship with their naturally spoken counterparts and thus increase sinewave vowel intelligibility.

In Hillenbrand, Clark, and Baer’s (2011) first experiment, 71 participants were recruited from an introductory phonetics class. They were first given a vowel intelligibility test using a set of 300 sinewave vowels drawn from the 1,668 /hVd/ syllable database collected by Hillenbrand, Getty, Clark, and Wheeler (1995). The syllables were recorded by 48 women, 45 men and 46 10 to 12 year old children and contained twelve vowel choices (ι,I,ε,E,Θ,A,□,o,Y,υ,℘,™).The majority of these speakers were from southern Michigan and surrounding Midwest states, as were the listeners of the Hillenbrand, Clark, and Baer (2011) study.

After the initial vowel intelligibility test, listeners were randomly assigned one of four training conditions. The first was a feedback test similar to the initial intelligibility test, except that the correct vowel category was indicated to the participant after they responded. The second was a sentence transcription task during which participants were asked to transcribe short sinewave sentences. The third condition was referred to as ‘triad’; participants in this condition listened to a sinewave vowel, followed by a naturally spoken version of the same vowel, and followed again by the sinewave version of the vowel. They received feedback after their response. The fourth condition was a control condition and involved an irrelevant task in which listeners were asked to judge whether naturally spoken /hVd/ syllables were spoken by men or women. All conditions were followed with the vowel intelligibility test again to assess the intelligibility rates post-training.

For the feedback and triad conditions, a set of 180 vowels was generated using the same database (Hillenbrand et al., 1995) but excluding those 300 already chosen for the initial intelligibility test. The test signals (the sinewave replicas of those recorded in the database) were created by extracting the frequencies of the formant peaks during the vowel. The sum of the three sinusoids corresponding to each peak in the signal created the sinewave replica test signal. For the sentence transcription task, 50 sentences were randomly chosen from the 250-sentence database Hearing in Noise Test (HINT) (Macleod & Summerfield, 1987; Nilsson, Soli, & Sullivan, 1994). These utterances were short, simple in structure and spoken by an adult male. An automated synthesis method was used to generate the sinewave sentence test signals from these recordings. The method was a broadband version of the method described in detail in the study by Hillenbrand, Clark, and Houde (2000).

Average sinewave vowel intelligibility across prior to training was 55.0%. This is roughly 40 percentage points below the average intelligibility of naturally spoken vowels (95.4%), and about 33 percentage points lower than the intelligibility of formant synthesized versions of the same vowels (88.5%) (Hillenbrand and Nearey, 1999). However, as the authors pointed out, 55.0% is still higher than the number of correct answers one might expect simply by chance (8.3%). This indicates that sinewave replicas of vowels are, in fact, conveying some phonetic information. Following the four training condition, participants' test scores of all conditions showed improvement, indicating to the authors that exposure to sinewave vowels alone increases intelligibility amongst listeners (Figure 2). The two conditions that showed the most improvement were the feedback and triad training conditions. Participants in the feedback condition improved by 7.5 percentage points, and participants in the triad condition improved by 10.9 percentage points. However, even with the highest improvement in intelligibility, participants still scored low when compared to identification of naturally spoken vowels (63.2%).

Hillenbrand, Clark, and Baer (2011) ran a second experiment to test the effects of more extensive training. The databases and sets of sinewave vowels used in the first experiment were also used in the second experiment. A separate group of 12 participants were recruited who had already completed the introductory phonetics class in which the other participants were enrolled. These participants were of the same geographic region of the speakers after whom the sinewave vowels were modeled. As in the first experiment, the vowel intelligibility test using the 300 sinewave vowel set was given before and after the training sessions were completed. The training condition in this experiment was a variation of the triad, most successful training in the previous experiment. In this condition, participants were presented with a sinewave vowel from the set of 180. After choosing a button on the computer screen, they were given feedback by blinking the correct button. If the participants chose correctly, they were moved on to the next trial. If they chose incorrectly, they were presented with the sinewave vowel, the naturally spoken corresponding vowel, and the sinewave vowel again. Participants in this condition showed an increase of 19.8 percentage points from the initial vowel intelligibility test to the post-test. While this number is higher than the training condition results seen in the first experiment, 74% intelligibility is still about 20 percentage points lower than intelligibility seen for naturally spoken vowels (Hillenbrand & Nearey, 1999).

The results of the two experiments performed in Hillenbrand, Clark, and Baer's (2011) study suggest that phonetic information can, in fact, be conveyed through isolated sinewave vowels, albeit with much lower accuracy than natural speech. When compared with the accuracy of the sentence transcriptions in the second condition (89.6%), the 74.0% vowel identification still falls short. In a subsequent study, Hillenbrand, Clark, Houde, K. Hillenbrand, and M. Hillenbrand (2012) set out to understand the discrepancy between sentence and vowel intelligibility. While it is obvious that listeners use higher level linguistic knowledge to reconstruct unfamiliar sentences, the authors proposed another possibility. The single-syllable test signals used in vowel intelligibility experiments are considerably shorter than the more extensively used sinewave sentences. They hypothesized that during longer utterances, listeners have the opportunity to make perceptual accommodations to the unusual acoustic characteristics of sinewave speech.

For this study, /hVd/ sinewave syllables were presented both in isolation and following a short carrier phrase, "The next word on the list is /hVd/." Ten men and ten women with an upper-Midwest dialect produced the carrier phrase with sixteen different vowels (ι,I,ε,E,Θ,A,□,o,Y,υ,℘,™,αY,αI,□I,φυ). The syllables and carrier phrases were then excised using a waveform editor in order to manipulate the utterances listeners were to hear. Using the same methods described in Hillenbrand, Clark, and Baer (2011), sinewave replicas of the signals were created. A total of 103 phonetically trained listeners with the same dialect as the speakers served as listeners. The listeners were randomly assigned to seven experimental conditions. Three of these conditions used naturally spoken syllables, and four used the sinewave replicas

(Table 1). In all conditions listeners were instructed to identify the vowel being spoken by choosing the corresponding button on a computer screen (Figure 6).

Results from Hillenbrand et al. (2012) showed consistent results across all natural speech conditions. In isolation, listeners identified the vowels with 95.1% accuracy, with 96.5% accuracy when preceded by a carrier phrase of the same speaker, and with 95.2% accuracy when the carrier phrase and syllable speakers were randomly paired. Of the sinewave speech conditions, vowel intelligibility was highest in the SWCP -WT condition (Figure 3). Though the vowel set was the same for all conditions, intelligibility was 73.1% in SWCP-WT compared to 49.1% when sinewave syllables were presented in isolation. Vowel intelligibility was still higher when speakers did not match in SWCP-XT, than the isolated syllable condition (SWIS) with intelligibility results of 61.3%. An analysis of the SWCP-XT condition showed that intelligibility increased when the sex of the carrier phrase speaker and vowel syllable speaker were randomly matched. Put differently, “the difference in average vowel intelligibility between SWCP-WT and different – sex trials from the corresponding XT condition is more than twice as large as that for same-sex trials (15.8 versus 6.7 percentage points)” (Hillenbrand et al., 2012, p. 7-8). This indicates that the lower intelligibility scores in the SWCP-XT condition are due primarily to listeners having greater difficulty identifying vowels when the carrier phrase was spoken by a woman and the test syllable by a man, or vice versa. In the NSCP-WT condition, intelligibility was 42.1%, the lowest of the four sinewave conditions.

The results of the Hillenbrand et al. (2012) study indicate that carrier phrase intelligibility results are due in part to speaker normalization, but not solely. In the SWCP-XT condition enhanced intelligibility scores were still produced in the presence of a carrier phrase with a different speaker than the following vowel. Furthermore, while intelligibility results were highest when carrier phrase and vowel syllable speakers matched, the NSCP-WT condition provides rich cues to speaker identity, yet scores were the lowest of all conditions. High intelligibility rates in SWCP-XT show that speaker normalization is not the only factor affecting intelligibility scores. The authors suggest that listeners also accommodate to the acoustic characteristics of the sinewave speech signal because they showed higher intelligibility than the SWIS condition even when speakers did not match. Much like in Remez et al.'s (1981) original study, where simply instructing listeners to hear speech improved intelligibility scores, the authors hypothesized that providing an example of sinewave speech encourages listeners to hear the test syllable as speech, resulting in an increase in intelligibility.

To further investigate this point, the authors ran a second experiment with sinewave vowels and carrier phrases. In this experiment, Hillenbrand et al. (2012) recruited a new set of 23 listeners. Using the same computer program (Hillenbrand and Gayvert, 2005), listeners were presented with alternating utterances from the SWIS and SWCP conditions in ten blocks of 32 trials (Figure 4). The results for this experiment, all within-subject, showed two prominent effects. First, listeners improved as they progressed through the experiment, a result also seen and discussed in Hillenbrand, Clark, and Baer (2011). Second, results alternate between high intelligibility scores in the presence of a carrier phrase and lower intelligibility scores in the presence of isolated syllables. In other words, as the carrier phrase came and went, so did the carrier phrase enhancement effect.

The results of these studies suggest that the presence of a sinewave carrier phrase is critical in signaling the brain to process the sounds with the same pattern-matching mechanisms as it does natural speech. The present study was designed to determine the importance of an intelligible carrier phrase. To test this, sinewave carrier phrases in an unfamiliar language will be compared with English carrier phrases. Japanese was chosen to compare to English because the Japanese phonetic inventory has no sounds which do not also occur in English, and listeners were very unlikely to have studied the language. The comparison of the English and Japanese carrier phrases allows us to determine more about the acoustic accommodation in carrier phrase enhancement seen in Hillenbrand et al. (2012). If the magnitude of the carrier phrase enhancement effect is the same for the English and Japanese conditions, it may suggest that listeners are attending more to the acoustic characteristics of the signal than the content of the utterance to aid in their response. If the magnitude of the enhancement is less for Japanese than English, it may suggest that listeners require an intelligible utterance prior to the test signal to identify that they are about to hear speech and process it as such.

Methods

Stimuli

For the test signals used in this study, one male speaker recorded the fixed carrier phrase "The next word on the list is…" followed by one of each of the 16 /hVd/ syllable choices (ι,I,ε,E,Θ,A,□,o,Y,υ,℘,™,αY,αI,□I,φυ). He recorded the same carrier phrase in Japanese. The same speaker also recorded 240 sentences from the 250 sentence Hearing in Noise Test (Macleod and Summerfield, 1987; Nilsson et al., 1994) in both English and Japanese. The sentences in this database are syntactically and semantically simple as well as short in duration (e.g. "He really scared his sister"). A waveform editor was used to excise the carrier phrases and /hVd/ syllables so that they could be mixed and randomly matched. The same automated sinusoid synthesizer described in Hillenbrand et al. (2011) was used to generate sinewave versions of all the recorded material. Briefly, the synthesizer works in 10-ms frames. As shown in Figure 5, for every 10-ms a 32-ms Hamming-windowed Fourier spectrum is generated as well as a masking threshold calculated as the 328 Hz Gaussian-weighted average of the spectral amplitudes in that spectrum. The masking threshold is then subtracted, point by point, from the Fourier spectrum so that valleys below the threshold are set to zero. Each amplitude in the spectrum is averaged over a 200 Hz band. The average is center-weighted so that the amplitudes near the middle of the 200 Hz band receive the most weight. This averaging removes the harmonics and leaves an envelope containing the averaged amplitudes. The peaks are extracted from the envelope, and sinusoids with the corresponding frequencies and amplitudes of those peaks are generated and summed to create the test signal. This synthesizer differs from the hand-edited signals used in the Remez et al. (1981) experiment, and all other sinewave speech studies, in that it does not track formants. The sinusoidal synthesizer finds all peaks in a frame (which most often correspond with the formants, though not necessarily). If a peak is continuous from frame to frame, the synthesizer will interpolate the peaks. If a peak is not continuous into the next frame, its amplitude will be ramped down to zero. If a new peak is found that is not continuous in frequency with any peak from the previous frame, than the amplitude is ramped up from zero to its measured amplitude over 10 ms to create this new peak. The result is a variable number of peaks that may start or end or continue between any given 10-ms frame. The test signals generated from this synthesizer, are quite intelligible and have the added benefit of being far less time-consuming to generate.

Participants

43 participants were recruited for this experiment. With ages ranging 19 to 50 years old, the majority of the listeners were from the same geographical region as the speaker with the same dialect. All participants were students at Western Michigan University who had completed a course in phonetics and were trained in phonetic transcription. Phonetically trained listeners were

chosen because research has shown that many errors made by untrained listeners are simply errors in transcription (Assmann, Nearey, & Hogan, 1982).

Procedures

Listeners participated in the self-paced experiment alone in a quiet room. Shown in Figure 6, they were presented with a practice task consisting of 32 trials of 16 different naturally spoken vowels presented in isolation and spoken by the same speaker. Listeners repeated any incorrect trials until they correctly identified all of the vowels. For the practice task, as well as every condition, they were allowed to replay the signal as many times as necessary before responding.

Listeners were then randomly assigned to one of six conditions, each consisting of 15 presentations of the 16 vowels in random order for a total of 240 trials. The conditions were: 1.) ISNS (n=3): Listeners were presented with naturally spoken recordings of the 16 vowels in /hVd/ syllables in isolation in random, 2.) ISSS (n=9): Listeners were presented with sinewave synthesized /hVd/ syllables in isolation, 3.) CP-E (n=9): Listeners were presented with the sinewave synthesized carrier phrase “The next word on the list is…” followed by one of the 16 /hVd/ sinewave syllables, 4.) CP-J (n=7): Listeners were presented with a Japanese sinewave synthesized carrier phrase which translates to the English phrase “The next word on the list is…” followed by one of the 16 /hVd/ sinewave syllables, 5.) HINT-E (n=8): Listeners were presented with one of 240 sinewave sentences from the Hearing in Noise Test database followed by one of the 16 /hVd/ sinewave syllables, and 6.) HINT-J (n=8): Listeners were presented with one of 240 Japanese sinewave sentences that translate to the same sentences used in HINT-E from the Hearing in Noise Test database followed by one of the 16 /hVd/ sinewave syllables. Prior to beginning their condition, the listeners were informed that what they were about to hear was unusual sounding speech. All carrier phrases and syllables were randomly shuffled prior to each listener beginning the experiment. Most listeners completed the pre-condition and assigned condition in 30-40 minutes.

Results

Mean intelligibility rates for all conditions in this experiment are shown in Figure 7. Consistent with previous studies, the control condition yielded a 99.6% intelligibility rate (Hillenbrand et al., 2011, & Hillenbrand et al., 2012). This result is not surprising and represents the high accuracy with which listeners identify naturally spoken speech sounds, especially when spoken carefully by a single talker. The ISSS condition's mean intelligibility was 52.4% and CP-E resulted in 77.3% accuracy. These results confirm the moderate intelligibility of isolated sinewave vowels as well as the improvement seen when those vowels are preceded by an intelligible carrier phrase (Hillenbrand et al., 2012). Both conditions using Japanese carrier phrases resulted in intelligibility rates almost identical to ISSS. With 50.1% and 48.6% accuracy rates for CPJAP and HINT-J, respectively, the results are almost as if the carrier phrase never occurred at all. Simply having time to listen to and adjust to the acoustical characteristics of sinewave speech did not help to increase intelligibility. The HINT-E condition resulted in a mean accuracy of 61% - about 15 percentage points below the fixed carrier phrase condition.

Discussion

The purpose of this study was to assess the carrier phrase enhancement effect on sinewave vowel intelligibility. The results indicate that the carrier phrase effect depends on the presence of an intelligible carrier phrase. In other words, an intelligible carrier phrase appears to be a key component in the unconscious interpretation of the sounds as speech. This is supported by the results of the Japanese carrier phrases, which were unintelligible and failed to increase intelligibility scores. The comparison of the HINT-E and CP-E conditions further supports this thinking. HINT-E, the condition with 240 different carrier phrases, resulted in intelligibility rates roughly 15 percentage points below the fixed carrier phrase condition. The condition containing a single, fixed carrier phrase repeated in all 240 trials was more intelligible to listeners than 240 different carrier phrases. The result of a more intelligible carrier phrase was higher intelligibility of the following syllable. This comparison indicates that intelligibility is an important part of unconsciously processing the utterance as speech and not noise, or music, or any other type of auditory input. However, when compared to the control condition, the most intelligible condition, CP-E, is still roughly 22 percentage points lower in intelligibility. From this we can determine that there are other factors contributing to the brain recognizing utterances as speech. While it is not acoustical accommodation to the signal, what those components might be, as well as the extent of importance of each component, is still up for debate. As discussed in Hillenbrand et al. (2011), speaker normalization is an important factor in addition to intelligibility. Future research may include changing the latency between carrier phrase and syllable presentation to assess whether temporal cues affect intelligibility rates. Experiments testing long, short, and inconsistent latency times may provide insight to the temporal relationship between test signals and its effects on intelligibility. Additionally, the experiment described in this paper should be repeated with more participants so as to gather statistically meaningful results.

Acknowledgements

This project could not have been possible without the help of Jeffery Angles of Western Michigan University’s Foreign Languages Department. In addition, I’m extremely grateful for the constant patience of Dr. Michael Clark, my thesis chairperson Dr. James Hillenbrand, and committee members Dr. Helen Sharp and Dr. Stephen Tasko of Western Michigan University’s Department of Speech-Language Pathology and Audiology.

References

Assmann, P., Nearey, T., and Hogan, J. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71*, 975-989.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5 I), 3099-3111.

Hillenbrand, J. M., Clark, M. J., & Baer, C. A. (2011). Perception of sinewave vowels. *Journal of the Acoustical Society of America, 129*(6), 3991-4000.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013-3022.

Hillenbrand, J. M., Clark, M. J., Hillenbrand, M. W., Hillenbrand, K. S., & Hourde, R. A. (2012) Perceptual accommodation to sinewave speech. In preparation.

Hillenbrand, J. M., Houde, R. A., & Gayvert, R. T. (2006). Speech perception based on spectral peaks versus spectral shape. *Journal of the Acoustical Society of America, 119*(6), 4041-4054.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 105*(6), 3509-3523.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-141.

Nilsson, M., Soli, S. & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 40-61.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947-950.

CONDITION NAME	SIGNALS PRESENTED
SWIS	Sinewave /hVd/ syllables in isolation.
ISNS	Naturally spoken /hVd/ syllables in isolation.
SWCP-SWIS-WT	Sinewave /hVd/ syllables preceded by a sinewave carrier phrase replicated from the same speaker.
NSCP – WT	Naturally spoken /hVd/ syllables preceded by a naturally spoken carrier phrase from the same speaker.
SWCP-SWIS-XT	Sinewave /hVd/ syllables preceded by a sinewave carrier phrase of a randomly paired speaker.
NSCP- XT	Naturally spoken /hVd/ syllables preceded by a naturally spoken carrier phrase of a randomly paired speaker.
NSCP-SWIS-WT	Sinewave /hVd/ syllables preceded by a naturally spoken carrier phrase of the same speaker.

Table 1. The seven experimental conditions of experiment 1 from Hillenbrand et al. (2012). Condition names are listed on the left and test signals presented are described on the right.

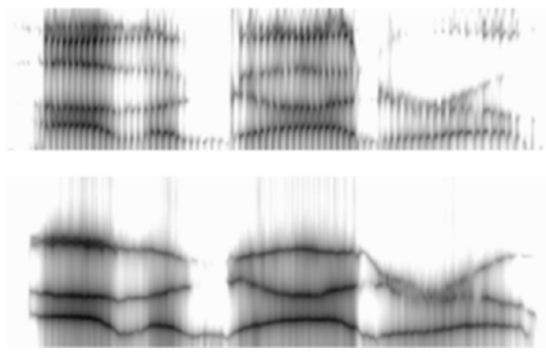

Figure 1. Top displays a spectrogram of the naturally spoken phrase “Avagadro”. The bottom is the sinewave replica of the same utterance.

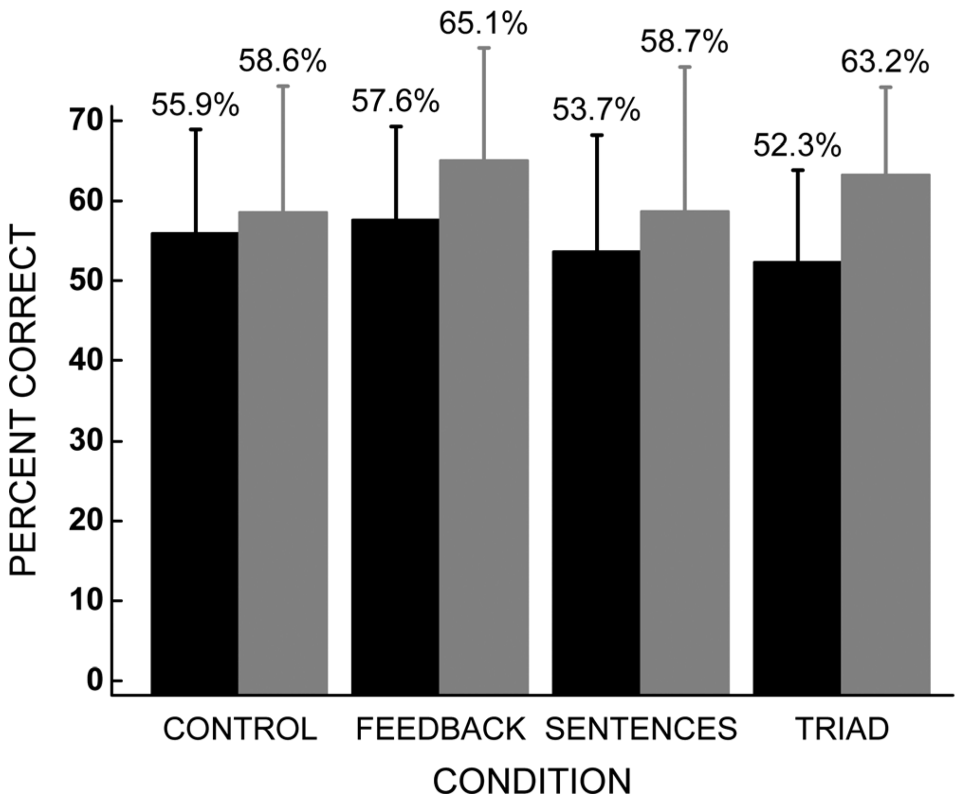

Figure 2. Results of experiment 1 from Hillenbrand, Clark, and Baer (2011). Dark bars show sinewave intelligibility of this initial vowel test and light bars show scores following the control and three training conditions. Error bars show one standard deviation.

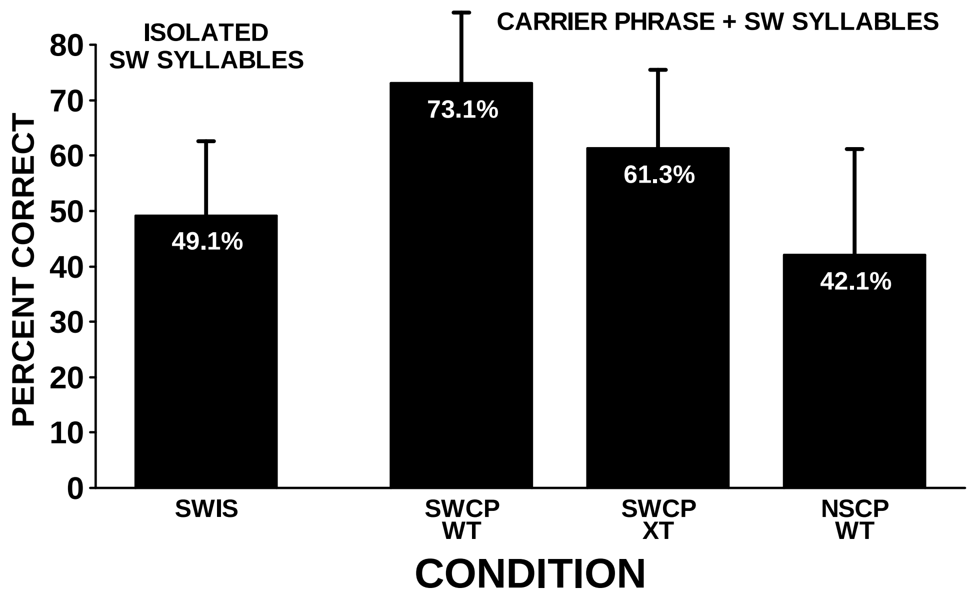

Figure 3. The results of experiment 1 from Hillenbrand et al. (2012). Error bars indicate one standard deviation. Conditions include SWIS – sinewave vowels presented in isolation, SWCP-WT – sinewave vowels preceded by a sinewave carrier phrase within talker, SWCP-XT- sinewave vowels preceded by a sinewave carrier phrase across talkers, NSCP-WT- sinewave vowels preceded by a naturally spoken carrier phrase.

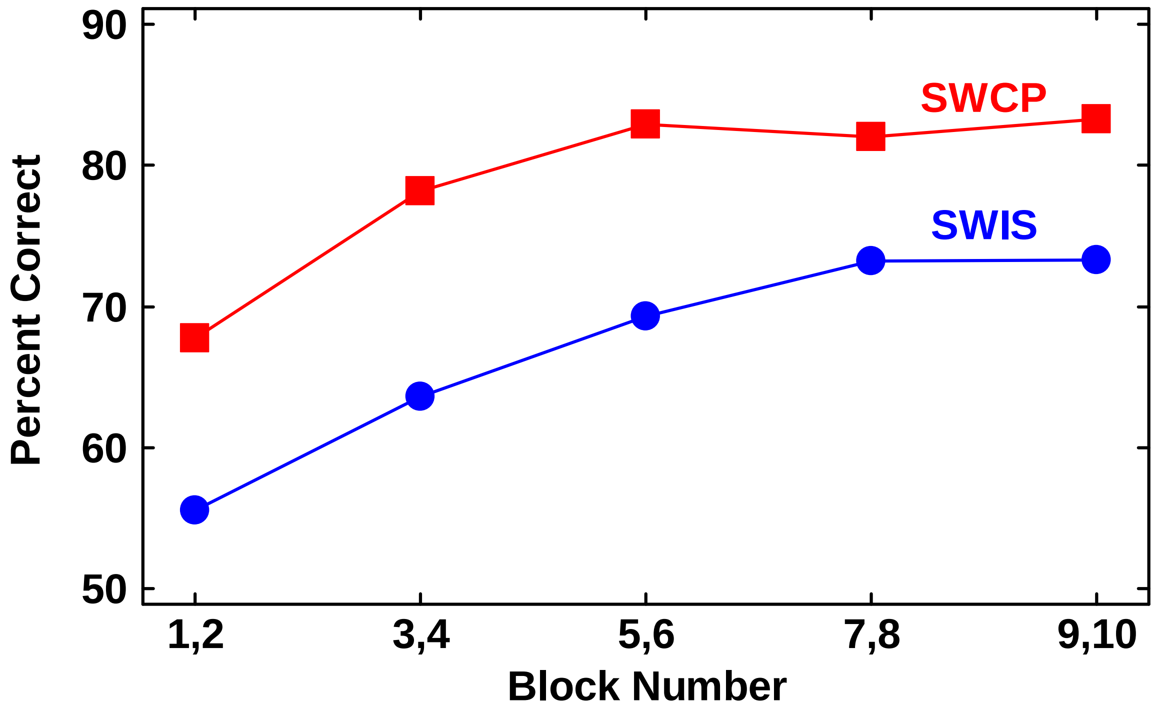

Figure 4. Results of experiment two from Hillenbrand et al. (2012). Vowel intelligibility scores of 32 blocks of alternating SWIS and SWCP test signals.

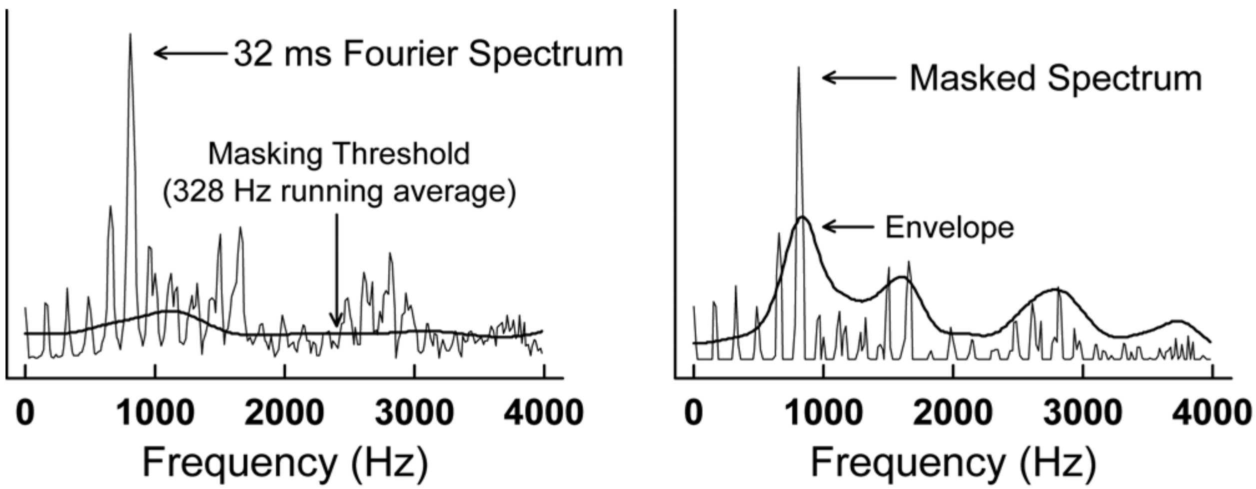

Figure 5. Steps used to process signals in the sinewave replica synthesis

Figure 6. Display of the computer screen looked like for all conditions in the experiment.

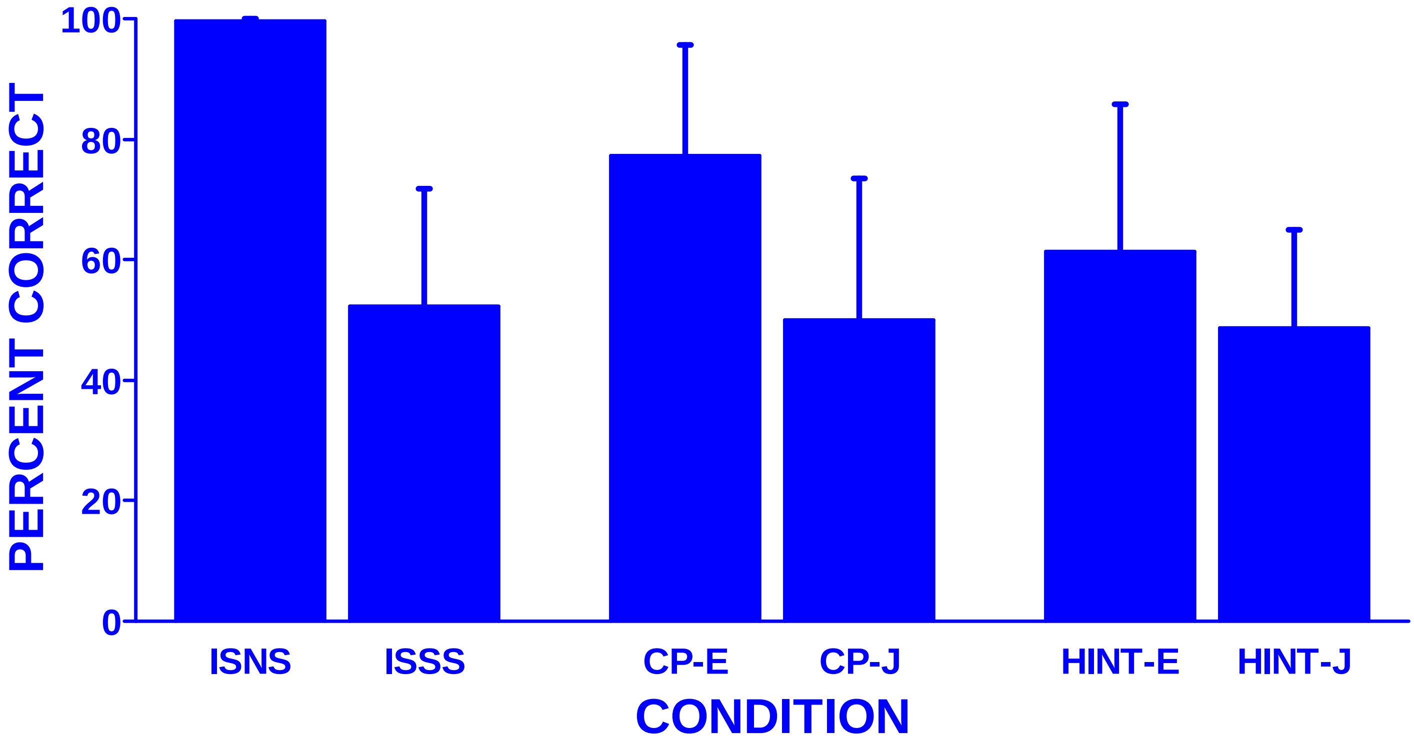

Figure 7. Vowel intelligibility results following English and Japanese carrier phrases. Error bars show one standard deviation. Conditions include ISNS – naturally spoken syllables presented in isolation, ISSS – sinewave syllables presented in isolation, CPENG – sinewave syllables preceded by a fixed English carrier phrase, CPJAP – sinewave syllables preceded by a fixed Japanese carrier phrase, HINT-E – sinewave syllables preceded by one of 240 English carrier phrases, and HINT-J – sinewave syllables preceded by one of 240 Japanese carrier phrases.