The Effects of Spectral Smearing on Hebrew Phoneme and Word Recognition

L. Kishon-Rabin, S. Patael, M. Menachemi, and N. Amir
Department of Communication Disorders, Sackler Faculty of Medicine, Tel-Aviv University, Israel
E-mail: lrabin@post.tau.ac.il, smadar_pa@hotmail.com, mayam10@hotmail.com, noama@post.tau.ac.il

ABSTRACT

The purpose of this study was to examine the effect of loss of spectral detail on speech perception in Hebrew. Spectral smearing was carried out by multiplying the speech signal by a series of low-passed white noise samples, causing tonal components in the signal to be replaced by noise. Smearing bandwidths of 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, 8000Hz and full bandwidth were used. Smearing was applied to 15 isophonemic lists, each with 10 one-syllable CVC Hebrew words. Results of eight normal-hearing subjects show that a (1) smearing bandwidth of approximately 1000Hz and 2000Hz reduced word and phoneme recognition to 50%, respectively, and (2) vowels were more susceptible to the effects of spectral smearing than consonants. These results are similar to the results reported in English and can be explained by the relative importance of spectral and temporal cues for the perception of speech.

1. INTRODUCTION

The general purpose of the present study was to determine how the recognition of Hebrew words and phonemes is affected by successive reductions of detail in the speech spectrum. Such information will allow us to (1) determine the relative importance of spectral information for the perception of different speech segments in Hebrew, and (2) to better understand the process of speech perception in Hebrew speaking hearing-impaired individuals.

Examining the relative importance of spectral information for the perception of Hebrew speech in subjects with normal hearing is important because the results are relatively independent of distortions in the temporal and intensity domains. It allows us to understand the limits of spectral distortion before intelligibility is affected. Such information can be implemented in speech transmission systems or auditory sensory devices that are limited in the amount of spectral information that they convey and/or transmit.

Boothroyd et al [1] measured the effects of loss of spectral detail in the acoustic signal on the perception of American English words by normal hearing listeners. By modulating the speech waveform by low-pass filtered noise (i.e., replacing each component of the speech spectrum by a band of noise whose bandwidth is twice that of the modulating noise), Boothroyd et al showed a decrease of phoneme and word recognition scores as the amount of spectral smearing bandwidth increased from zero to complete. Smearing bandwidths of 1400Hz and 720Hz were required to reduce phoneme and word recognition score to 50%, respectively.

While the effect of loss of spectral detail on speech perception is available for American English, this data cannot be implemented directly to a language, such as Hebrew that has a different sound system than English. While Modern Hebrew is limited to five vowels and 19 consonants, a simplified view of the American English phonemic inventory includes 12 basic vowels (not including diphthongs) and 24 consonants [2]. The two languages share 16 consonant phonemes /p, b, f, v, m, t, d, s, z, l, n, ʃ, j, k, g, h/ (with slight differences in pronunciation: e.g., Hebrew apical consonants are usually apical-dental while English apical consonants are usually apical-alveolar) [3]. In addition Hebrew has the consonants /ts, x, ɣ/ while English has the consonants /w, θ, ð, tʃ, dj, r, j, ng/. Although the consonants in Hebrew and American English differ, the same features classify them all: voicing, place of articulation, and, manner. The Hebrew vowel system does not include the tense-lax distinction as a distinctive feature, nor does it have low front, low back, or central vowels.

In terms of the acoustics of the speech sounds, data in Hebrew are available for vowel formants and voice-onset-time (VOT) of stop-plosives. Hebrew vowel space is much more centralized and more triangle in shape compared to that of English and that of languages with the same number of vowels represented by the same phonetic symbols (e.g., Spanish and Shona) [4]. The reduced vowel space implies smaller acoustic difference between the vowels in Hebrew compared to other languages. In terms of voiced plosives, although English and Hebrew use the same voice-voiceless distinction and the same phonemes (/p, b, t, d, k, g/), they appear to differ in their articulatory-acoustic patterns. Hebrew speakers use a considerable amount of voicing lead category for the voiced stop-plosives (VOT=-90msec), but use an in-between category (between the short-lag and long-lag) for the voiceless ones (VOT= +28 to +56 ms) [5]. In contrast, American English speakers typically use zero-onset/short lag (VOT= 0 to +3msec) for voiced stop-plosives and long lag (VOT=+50 msec or more) for the voiceless ones [6].

ISBN 1-876346-48-5 © 2003 UAB

Thus, the main purpose of the present study was to examine whether the acoustics of the Hebrew language is affected differently (from English) by the loss of spectral detail. The present study used the same technique as Boothroyd et al [1], namely the modulation of the speech waveform by low-pass filtered noise. Such information will also provide us with some insight to the speech process of Hebrew speaking hearing-impaired individuals. Reduced spectral information is only one of several effects of sensorineural damage whose extent is correlated with degree of hearing loss [7]. Loss of spectral detail is assumed to simulate, at least to a certain extent, the effect of widened auditory filters encountered in certain types of hearing loss [1, 7]. However, due to the many additional factors that coexist and interact within hearing-impaired individuals, one should consider the information obtained here limited in its contribution to understanding the relative importance of the many psychoacoustic consequences of sensorineural hearing loss.

2. METHOD

Subjects: Subjects were 8 female volunteers, 22-27years of age. All were native Hebrew speakers, with no history of hearing disorders and were confirmed to have a pure-tone thresholds between 0 and 15dB HTL, at octave intervals from 250 to 4000 Hz, bilaterally.

Stimuli: Stimuli consisted of the HAB word lists which are the Hebrew version of the AB word lists [8]. The HAB consists of 15 isophonemic lists with 10 consonant-Vowel-Consonant (CVC) meaningful words in each list. Stimuli were recorded in a sound-treated room by a Hebrew speaking young female native of Israel onto a computer at a sampling rate of 22.5 kHz using 16 bits per sample. Performance was measured as the percentage of phonemes and words correctly recognized.

Spectral Smearing: Spectral smearing was performed digitally. The smearing was carried out using a procedure similar to the one described by Boothroyd et al [1], in which the signals are multiplied by bandlimited noise. The rationale underlying this procedure is that multiplication in the time domain is equivalent to convolution in the frequency domain. If we consider a harmonic signal, its theoretical spectrum is actually a line spectrum. Thus, convolution with bandlimited noise in the frequency domain creates replicas of the noise spectrum centered at the harmonic frequencies.

The processing and filter design were carried out using Matlab software. The noise signals in were created by passing digital gaussian white noise through a set of low pass filters with successively wider bandwidths. These were order 200 FIR filters, designed using the window method with a Hamming window. The cutoff frequencies of these filters started at 125 Hz., up to 4000 Hz in octave increments. These filters have nearly negligible ripple in the passband (< 0.05dB), -6dB of attenuation at cutoff, stop band attenuation of at least 50 dB, and a stop band edge of approximately 150 Hz above the cutoff frequency. A total of six filters were constructed resulting in six different smearing bandwidths of 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, and 8000Hz.

Procedure: Stimuli were presented through the speech input of a Grason Stadler GSI-61 audiometer, and presented binaurally using Telephonics TDH-50P headphones. Subjects were seated in a sound-treated booth. Presentation was computer controlled, with a pause of 2.5 seconds between words. The recordings were normalized to have identical peak-to-peak amplitude, and were then presented at 60dB HL. Subjects were instructed to verbally repeat the stimuli and also to write it down. They were instructed to guess an answer in case they were not sure of what they heard.

Before starting formal testing, each subject received a training list with a spectral smearing bandwidth of 250 Hz. Testing consisted of 7 lists with increasing smearing bandwidths of 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, 8000Hz, which were followed by 7 lists with decreasing smearing bandwidths. At the end of testing, a list with no spectral smearing was administered. Thus, subjects listened to a total of 160 words. A Latin Matrix design was used.

3. RESULTS

The percent of words and phonemes correctly recognized was calculated for each subject and spectral smearing bandwidth. The mean group word data as a function of spectral smearing bandwidth are shown in Figure 1. Note that in order to compare between the present data in Hebrew and the existing data in English, the effect of spectral smearing on speech perception in American English, as reported by Boothroyd et al [1] are added onto each of the following graphs.

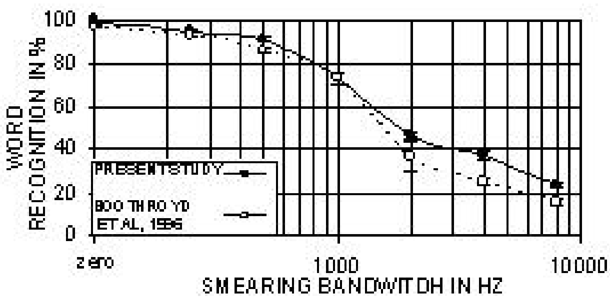

FIGURE 1: Mean Hebrew word recognition scores (in %) for isolated consonant-vowel-consonant words as a function of the amount of spectral smearing. Data points are means (±1 s.e.) for eight normally hearing, Hebrew-speaking adults. Shown for comparison are data from English-speaking adults [1].

Figure 1 shows that the average word recognition scores ranged between 98.75% for no smearing and 0% for full-band smearing. Also, performance decreased monotonically as the amount of spectral smearing

ISBN 1-876346-48-5 © 2003 UAB

increased. Specifically, mean group performance decreased to 86.25%, 78.75%, 54.38%, 16.25%, 10%, and 5%, for spectral smearing bandwidths of 250 through 8000 Hz, respectively. The decrease in performance between each two successive test conditions was found to be statistically significant ($p<0.05$) with the exception of 250-500Hz. The largest decrease in performance between two successive conditions occurred as the spectral smearing bandwidth changed from 1000 Hz to 2000 Hz. Note that 50% percent word recognition was obtained at spectral smearing bandwidth of approximately 1000Hz.

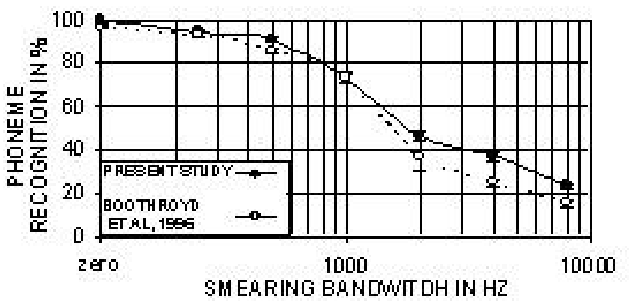

FIGURE 2: Hebrew phoneme recognition in isolated consonant-vowel-consonant words as a function of the amount of spectral smearing. See caption to Figure 1 for details.

The mean group phoneme recognition scores as a function of smearing bandwidth are shown in Figure 2. Performance decreased monotonically from 99.59% for no smearing to 95%, 91%, 73%, 46.65%, 37.29%, 23.33%, and 0 %, for spectral smearing bandwidths of 250 through 8000 Hz and full-bandwidth smearing, respectively. Again, the decrease in performance between each two successive test conditions was found to be statistically significant ($p<0.05$) with the exception of 250-500Hz. Fifty percent recognition of phonemes occurred at spectral smearing bandwidth of approximately 2000 Hz.

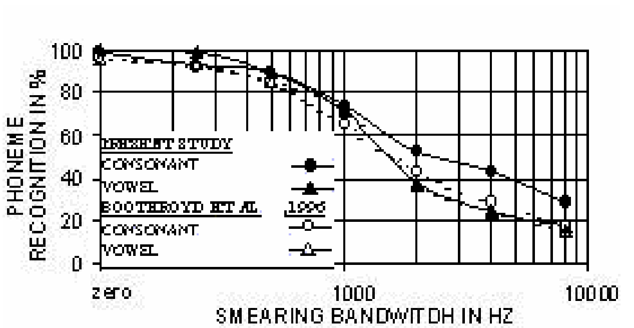

FIGURE 3: Hebrew vowel and consonant recognition scores (in %) as functions of the amount of spectral smearing. See caption to Figure 1 for details.

Figure 3 shows mean percent recognition scores for vowels and consonants as a function of spectral smearing bandwidth. Similar to words and phonemes, the recognition of both vowels and consonants decreased monotonically as the spectral smearing bandwidth increased. They both appear to have similar recognition scores for spectral smearing bandwidths of 250, 500 and 1000 Hz. Furthermore, for vowels and consonants, the greatest decrease in performance occurred between 1000 and 2000 Hz. However, for spectral smearing bandwidths equal or greater than 2000 Hz, the recognition of consonants was consistently better than that of vowels by approximately 20 percentage points. In other words, spectral smearing appeared to have a lesser effect on the recognition of consonants than that of vowels in Hebrew. Interestingly, 50% recognition of consonants occurred at spectral smearing bandwidth of approximately 2000 Hz whereas for vowels it occurred between 1000 and 2000 Hz bandwidth (Note that at spectral smearing bandwidth of 2000 Hz, mean group vowel recognition scores were reduced to 37%). It should also be noted that no statistically significant differences were found between consonants in initial and final position.

4. DISCUSSION

The finding of the present study that spectral smearing of 250 Hz bandwidth degrades speech intelligibility is in keeping with the hypothesis that spectral smearing becomes effective when the smeared bandwidth is of the same order of magnitude as that of the listener's auditory filters [1, 9]. A bandwidth of 250 Hz approximates the normal auditory bandwidth at 1500-2000 Hz [10]. Our findings in Hebrew, that increase in bandwidth beyond 250 Hz degrades speech recognition in a similar manner as in English supports the importance of this frequency region (of 1500-2000 Hz) in Hebrew as well. An alternative explanation proposed by Boothroyd et al [1] is that spectral smearing begins to have an effect when it significantly increases formant bandwidth, which is also in the region of 250 Hz. This would then support the notion that the formant pattern is the most important information-bearing aspect of the spectral envelope of speech.

In keeping with the English data, our Hebrew findings also show that smearing had a greater effect on word recognition than it did on phoneme recognition. It appears that for word recognition there is almost a complete effect of smearing at 2000 Hz bandwidth. In contrast, approximately 25% of the phonemes can be correctly identified at spectral smearing bandwidth of 8000 Hz. These differences may be attributable to the non-linear relationship between the two measures (phonemes and words) [1]. It is also possible, however, that these results reflect the unique interaction between the specifics of the Hebrew sound system and the acoustics of the language. Hebrew has only 5 vowels, located on a very reduced vowel space, thus emphasizing the importance of small spectral changes. Because whole word recognition depends also on the correct identification of vowels, and, vowels are highly dependent on spectral information, one would expect word recognition to drop in performance with loss of spectral detail. Thus the present studies findings are in keeping with this hypothesis.

The finding that consonant perception was less affected by smearing than vowel perception is in keeping with the results found in other languages [1, 9]. These findings can

 ISBN 1-876346-48-5

be explained in terms of the relative contributions of spectral and temporal cues to the perception of consonants and vowels. As indicated above, the small number of Hebrew vowels in a reduced vowel space requires relatively fine spectral resolution to perceive these vowels correctly. Loss of spectral details beyond smearing bandwidth of 250 Hz has an immediate effect on vowel perception. However, the ratio of the contributions of temporal and spectral cues is much higher in consonants. This is supported by our findings that a 1/3 of the consonants can be correctly perceived at smearing bandwidth of 8000 Hz compared to a ¼ of the vowels perceived at smearing of 4000 Hz. This data is also in keeping with other findings showing Hebrew consonant perception more resistant to loss of spectral information compared to Hebrew vowels [11]. In general, the present findings suggest that Hebrew is somewhat less susceptible to spectral smearing than English. However, more research is required before such conclusions can be drawn.

Our present findings do not show differences between consonants in initial versus final position. This is in contrast to the findings in English [1]. It may suggest that the temporal cues to the perception of Hebrew consonants may be different than those for English. For example, we already know that the final voicing in Hebrew is difficult to perceive based on temporal cues (e.g., lengthening of the vowel that precedes the final voiced consonant) [11,12]. Also, Hebrew listeners perceive initial voicing at shorter VOT values than English listeners [13]. Clearly, information regarding the acoustics of the Hebrew speech sounds and their relative importance in auditory perception will help further understand the differential response of the effect of spectral smearing on the perception of English and Hebrew speech.

5. CONCLUSIONS

The effects of spectral smearing bandwidth on phoneme and word recognition in Hebrew were similar to those reported in English [1] despite the differences in the speech sounds of the two languages. Recognition of words is more susceptible to the effect of smearing than are phonemes. While most words cannot be recognized at smearing level of 2000 Hz, a ¼ of the phonemes can be still identifiable at smearing of 8000 Hz. Consonants were found to be more resistant to the effects of spectral smearing than vowels. In contrast to English, no differences were found between initial and final consonants. Most of the data can be explained in terms of the relative importance of spectral versus temporal cues to the perception of Hebrew speech.

REFERENCES

[1] A. Boothroyd, B. Mulhearn, J. Gong and J. Ostroff, “Effects of spectral smearing on phoneme and word recognition,” *Journal of the Acoustical Society of America,* vol.100, pp.1807-1818, 1996.

[2] A. Boothroyd, *Speech Acoustics and Perception,* Austin TX: Pro-Ed, 1986.

[3] Y. Tobin, *Phonology as human behavior: Theoretical implications and clinical applications,* Durham: Duke U Press, 1997.

[4] T. Most, O. Amir and Y. Tobin, “The Hebrew vowel system: Raw and normalized acoustic data,” *Language and Speech,* vol. 43, pp. 295-308, 2000.

[5] L.J. Raphael, Y. Tobin, A. Faber, T. Most, B.H. Kollia, D. Milstein, “Intermediate values of voice onset time,” in *Producing speech: Contemporary issues for Katherine Safford Harris,* F. Bell-Berti and L.J. Raphael, Eds., AIP, 1995.

[6] G.J. Borden, K.S. Harris and L.J. Raphael, *Speech Science Primer,* Baltimore, MD: Williams & Wilkins, 1994.

[7] B.C.J. Moore and B.R. Glasberg, “Comparison of frequency selectivity in simultaneous and forward masking,” *Journal of the Acoustical Society of America,* vol. 80, pp. 93-107, 1986.

[8] A. Boothroyd, “Statistical theory of the speech discrimination score,” *Journal of the Acoustical Society of America,* Vol. 43, pp. 362-376, 1968.

[9] M. ter Keurs, J.M. Festen and R. Plomp, “Effect of spectral envelope smearing on speech perception,” *Journal Acoustical Society of America,* vol. 91, pp. 2872-2880, 1992.

[10] B. Scharf, “Critical Bands” in *Foundations of Modern Auditory Theory,* J.V. Tobias Ed, New York: Academic, 1970.

[11] L. Kishon-Rabin and M. Nir-Dankner, “The perception of phonologically significant contrasts using speech envelope cues,” J*ournal of Basic Clinical Physiology and Pharmacology,* Vol. 10, pp. 209-219, 1999.

[12] L. Kishon-Rabin and Y. Henkin, “Age related changes in the visual perception of phonological contrasts, ” *The British Journal of Audiology,* vol. 34, pp. 363-374, 2000.

[13] L. Kishon-Rabin, S. Rotshtein and R. Taitelbaum, “Underlying mechanisim for categorical perception: Tone-onset time and voise-onset time evidence of Hebrew voicing,” *Journal of Basic Clinical Physiology and Pharmacology*, Vol. 13, pp. 117-134, 2002.

ISBN 1-876346-48-5 © 2003 UAB