Appendix

Acoustic Analyses

Acoustic analyses were performed using two programs: Praat, version 4.1.2 (Boersme & Weenink, 2003) and the MDVP (Multi Dimensional Voice Program), model 5105, version 2.1 (Kay Pentax 2007, NJ). Specifically, the Praat program was used to measure acoustic parameters related to prosody: Articulation Rate, Mean Fundamental Frequency (mF0), F0-Range, and Amplitude-Range. The MDVP was used to measure two parameters related to voice quality: Period Perturbation Quotient – PPQ and Amplitude Perturbation Quotient – APQ. Following is a brief explanation of the acoustic parameters.

Articulation Rate

Speaking rate is defined as the number of spoken units (e.g., words/syllables) per unit of time (minute/second). It is calculated across continuous speech segments, which may include pauses, disruptions or dysfluency (Howell, Au-Yeung & Pilgrim, 1999). As such, speaking rate is considered a global measure of verbal output and language proficiency (Costello & Ingham, 1984). In contrast, *Articulation rate*, which was measured in the present study, is based only on fluent utterances, excluding pauses and dysfluency (Hall, Amir & Yairi, 1999; Howell, et al., 1999). As such, calculating articulation rate reduces linguistic effects, in comparison to overall speaking rate, and is mainly viewed as a measure representing articulatory motor control (Walker et al., 1992).

Figure 1, for example, presents a spectrogram and a time-wave display of a single sentence that was included in this study, as it appears in the Praat "edit" window. The utterance produced in this example was /bikaʃti bɛofɛn mɛfoɤaʃ kɛsɛf/

(which, in Hebrew, means "I specifically asked for money"). The upper half of the figure is the time-wave display, in which the horizontal axis represents time (in seconds), and the vertical axis represents signal intensity/amplitude (on a normalized scale), which is typically perceived as loudness. The lower half of the figure is a spectrogram, in which the horizontal axis represents time (in seconds), the vertical axis represents speech spectral frequency (in Hertz), and signal intensity is represented by the gray level of the image (darkness/density of print).

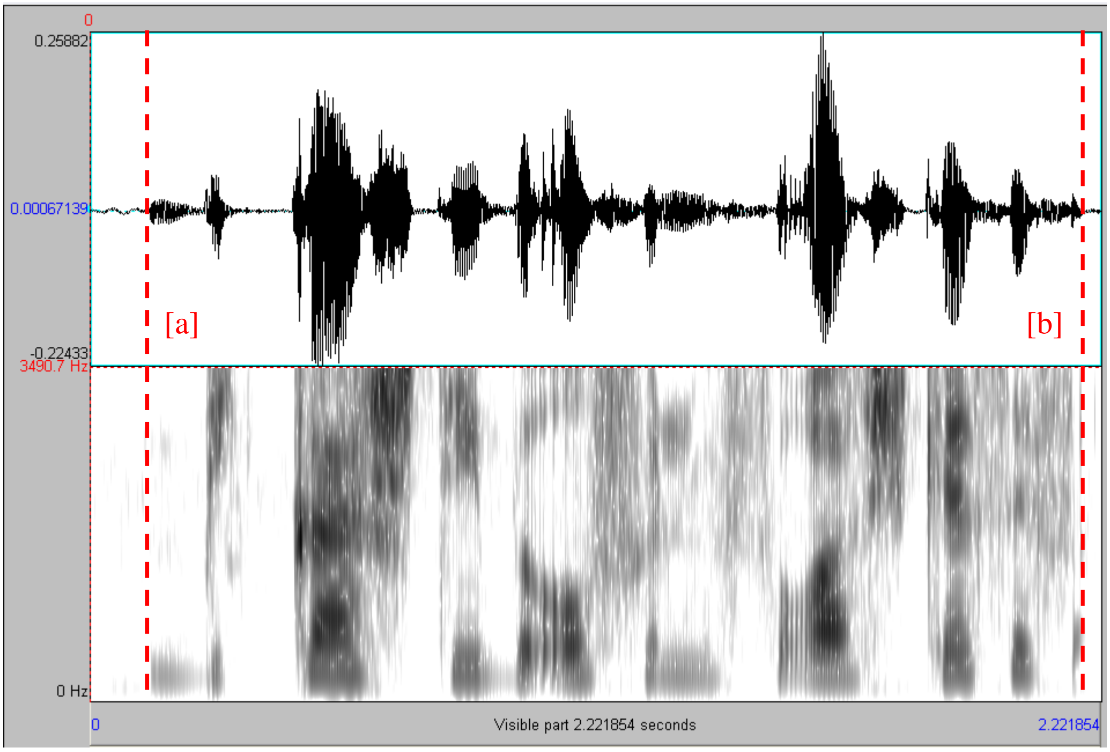

Figure 1. A combined time-wave display and spectrographic display of the utterance /bikaʃti bɛofɛn mɛfoɣaʃ kɛsɛf/. Utterance boundaries are marked with the vertical dashed lines [a] and [b].

The duration of the utterance is calculated by subtracting point [a] in time from point [b] in time. In this case, the duration of the utterance was 2.0408 sec. This particular utterance consisted of four words; therefore articulation rate was calculated as [4 words / 2.0408 sec. X 60] = 117.60 words per minute (WPM).

Mean Fundamental Frequency (mF0)

The fundamental frequency of the voice represents the vibratory rate of the vocal folds during the production of voice (phonation) and speech, and is quantified in Hertz (Hz). It is closely correlated, though not linearly, with perceived pitch (Stevense & Volkmann, 1940). Fundamental frequency (F0) is measured by identifying a cyclic (periodic) pattern in the acoustic signal, measuring the duration of each cycle, and estimating how many cycles can be produced per second. One can only measure fundamental frequency in the voiced segments of a spoken utterance (when vocal folds are vibrating), and it is most reliably measured during the pronunciation of vowels.

Figure 2 presents a short segment of 34.939 msec, extracted from the vowel /a/ in the word /bikaʃti/, appearing in the sentence presented in Figure 1. This figure was extracted directly from the Pratt "edit" window. The vertical blue lines appearing on the time-wave display (upper half of the figure) mark the boundaries between each individual voicing (periodic) cycle. This measurement represents the fundamental frequency for that moment in time. By averaging these values across a selected segment, a mean fundamental frequency can be calculated. In Praat, this procedure can be performed using the "voice report" command. For example, the mean fundamental frequency (*mF0)* for the short segment in Figure 2 was 216.77 Hz. Overall *mF0* for the complete utterance, from which this segment was extracted, was 187.48 Hz. These values are within normal range for women in general (Baken, 1987), and in Hebrew in particular (Most, Amir & Tobin, 2000).

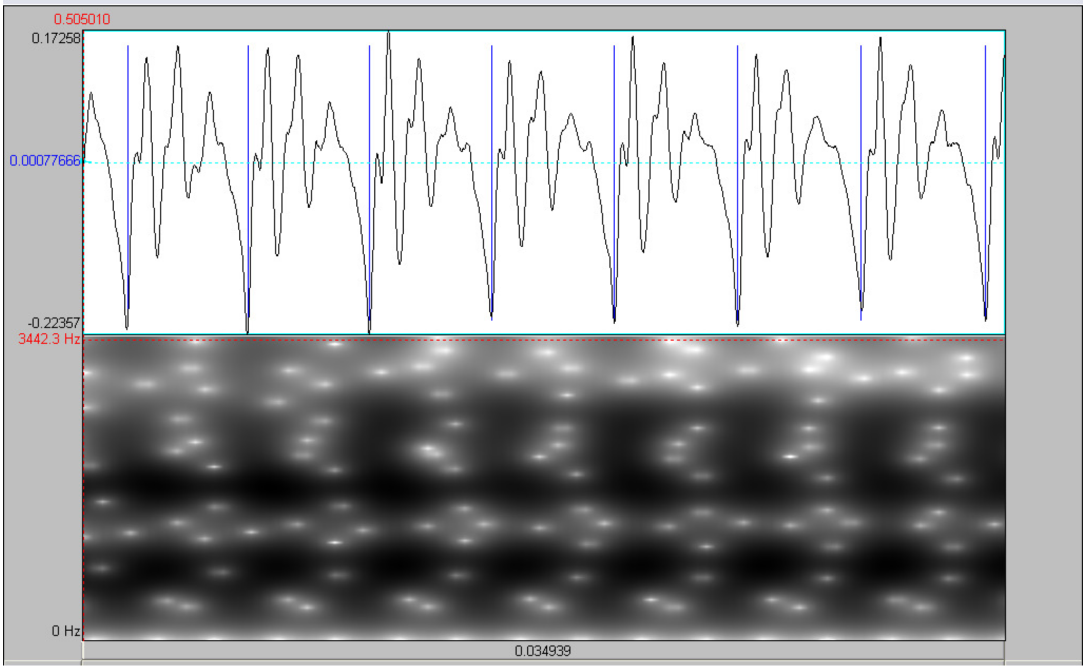

Figure 2. A combined time-wave and spectrographic display of a short segment of the vowel /a/, extracted from the word /bikaʃti/, using the Praat program. Blue vertical lines on the time-wave display mark individual voicing cycle boundaries.

Fundamental Frequency Range (F0-Range)

As described above, fundamental frequency can be measured separately, based on each vibratory cycle. Therefore, fundamental frequency can vary across a given speech segment, depending on changes in the speech pattern. These changes in fundamental frequency are perceived by the listener as changes in pitch, and are typically referred to as *intonation*. *F0-range* is calculated by subtracting the minimum F0 value observed in a given segments from the maximum F0 value.

To perform this measurement with Praat, it is necessary to track changes in the fundamental frequency across the selected speech segment. Figure 3 illustrates an example of an analysis of the sentence /vɛgam bɛʔɛzɛ ʃɛhu makom ʔaxʃav/. The blue curved line on the spectrogram represents a continuous measure of F0 over time. Note that the curved line tracks F0 only in the voiced segments on the spoken utterance. Therefore, breaks in this line represent unvoiced or silent segment, where F0 could not be traced. F0 analysis display range was set, in this Figure, to 100-300 Hz, to match expected F0 values for women. By placing a cursor on the highest peak of this curved line, a measurement of the highest F0 value in the speech segment is obtained (point [a]). Similarly, by placing a cursor on the lowest point of this curved line, the lowest F0 value is obtained (point [b]).

In this example, the highest F0 value was 211.32 Hz, whereas the lowest F0 value was 146.56 Hz. Therefore, *F0-range* for this utterance was calculated to be 54.76 Hz.

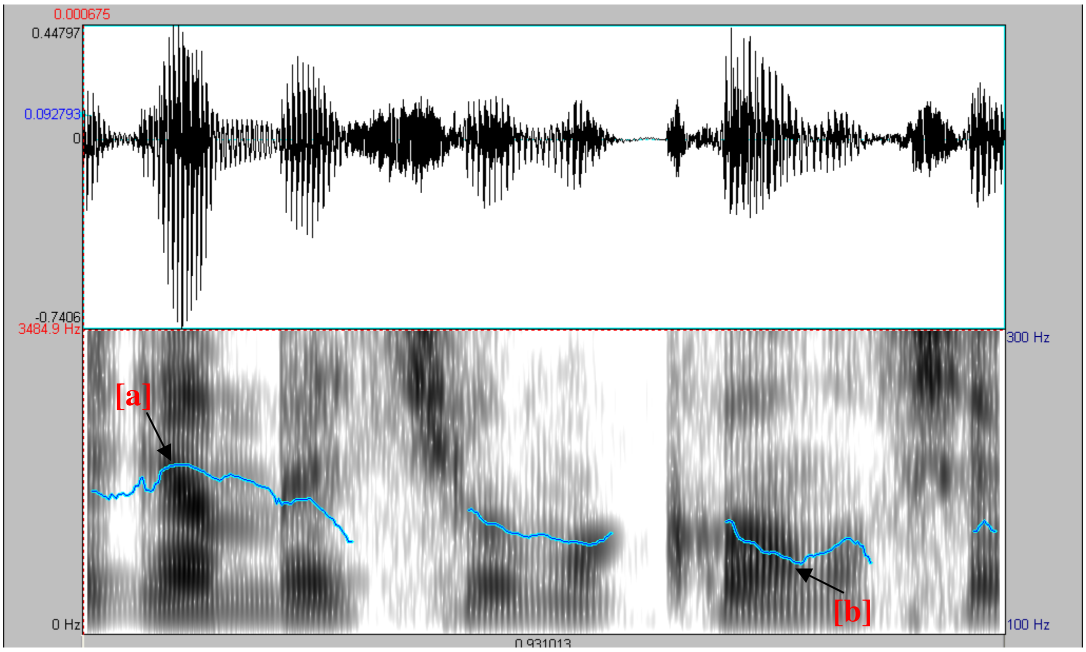

Figure 3. An illustration of tracking of the fundamental frequency (F0), using the Praat "edit" display, across a speech segment. Point [a] represents the highest F0 value, and point [b] represents the lowest value.

Amplitude range

Amplitude range is a measure of the amount of change in intensity (amplitude) within the speech segment, and it is calculated in decibels (dB). It is closely correlated with the subjective perception of loudness, though not linearly (Fletcher & Munson, 1933), and as such it is part of the *prosodic* characteristics of speech (together with rate of speech/articulation and F0 dynamics). This measurement is performed, much like the measurement of the F0-range, based on the identification of discrete vibratory cycles. The peak amplitude for each vocal cycle is identified first. Then, the minimum amplitude value identified across the selected speech segment is subtracted from the maximum value.

Figure 4 illustrates an example of an analysis of the sentence presented in Figure 3. The yellow line on the spectrogram represents a continuous measure of amplitude over time. Peak amplitude is revealed by placing the cursor on the highest point of the curved line (point [a]). Similarly, by placing a cursor on the lowest peak of this curved line, the lowest amplitude value is revealed (point [b]). In this example, the highest amplitude value was 80.94 dB, whereas the lowest amplitude value was 68.43 dB. Therefore, *Amplitude-range* for this utterance was calculated to be 12.51dB.

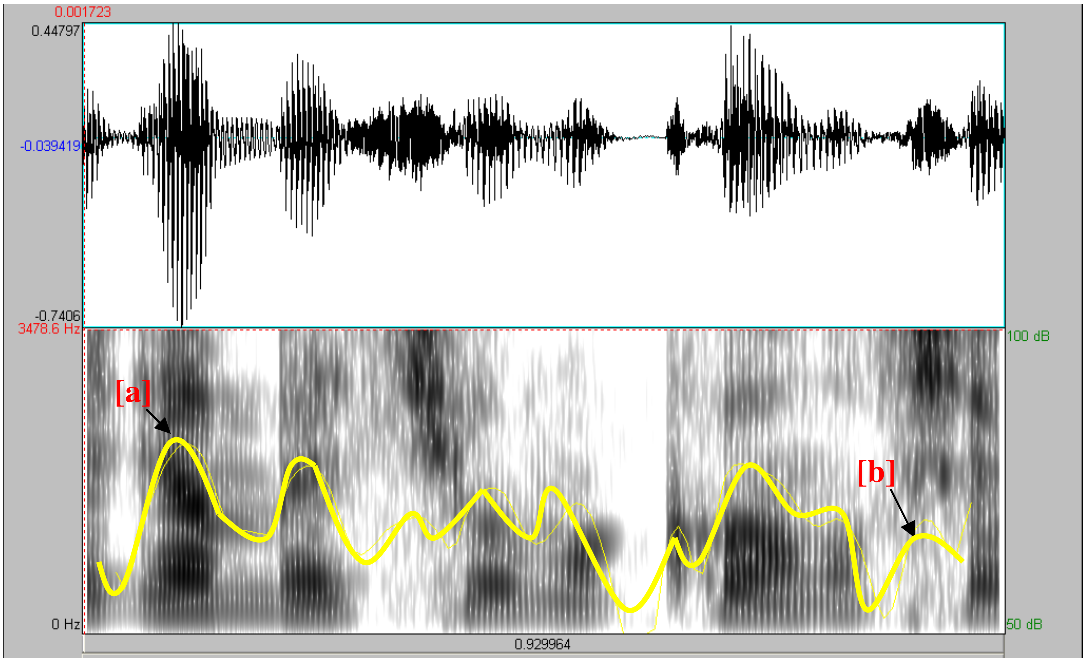

Figure 4. An illustration of tracking of intensity (amplitude), using the Praat "edit" window, across a speech segment. Point [a] represents the highest amplitude value, and point [b] represents the lowest value.

Frequency Perturbation

Frequency perturbation, typically called *jitter*, refers to immediate (short-term) variations of the fundamental frequency (Baken, 1987). During a sustained phonation of a steady vowel, although the speaker intends to maintain a steady pitch level, the duration of each individual vocal cycle is slightly different from that of the cycle that follows it. These differences are termed jitter.

Figure 5 illustrates an example of a short segment of 38.72 msec extracted from a sustained phonation of the vowel /a/, produced by a male speaker. The vertical blue lines on the time-wave display represent the start and finish of the individual vocal cycles. The differences in F0 values for each vibratory cycle (printed between the markers) illustrate mild perturbations. Jitter is calculated as the mean difference (in percentage) between each successive pair of cycles. In this case, jitter value for this short segment was calculated to be 0.39%.

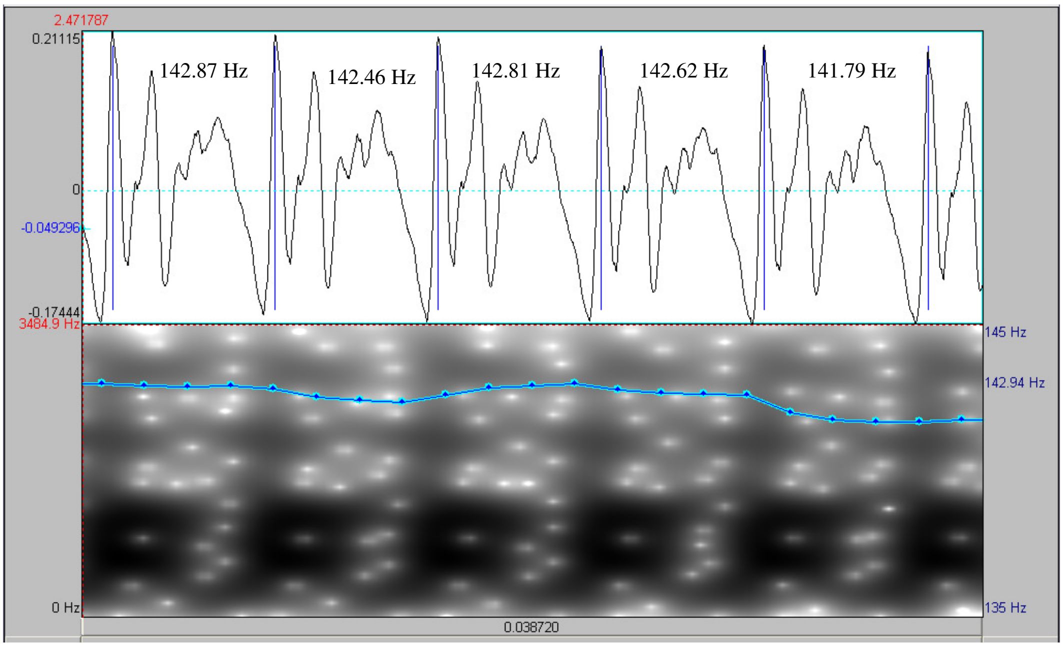

Figure 5. An illustration of a short segment extracted from a sustained phonation of the vowel /a/, produced by a male speaker. Vertical blue lines on the time-wave display represent start and end of individual phonatory cycles, and F0 measurements for each cycle are presented.

Jitter is considered most appropriate for use in sustained phonations of vowels, since it does not account for voluntary changes in F0. However, when continuous speech is examined and voluntary changes in F0 appear (i.e., intonation), the use of perturbation measures based on a smoothing factor is considered preferable (Baken, 1987). *Pitch Period Perturbation Quotient (PPQ)* is an acceptable measure for this purpose. It calculates a relative evaluation of the period-to-period variability of the fundamental frequency within the analyzed signal with a smoothing factor of five cycles. Specifically, it averages the differences between the values of a middle cycle to those of two preceding cycles and two successive cycles. Then, the same calculation is repeated for all series of five adjacent cycles using a moving window. In the present study, PPQ measurements were performed using the MDVP.

Amplitude perturbation

Amplitude perturbation, typically called *shimmer*, refers to immediate (short-term) variations of amplitude (Baken, 1987). In a manner similar to jitter, during a sustained phonation of a vowel, the maximum amplitude of each individual vocal cycle is slightly different from that of the cycle that follows it, although a steady production is intended by the speaker.

Figure 6 uses the same example displayed in Figure 5 to demonstrate amplitude differences between adjacent cycles. The maximum amplitude value obtained for each vibratory cycle (in this case, it was identified at the vertical blue markers) is printed next to the vocal markers. The small differences in intensity between the adjacent cycles represent *shimmer*. It is calculated as the mean difference of amplitude (in percentage) between each successive pair of cycles. In this example, the shimmer value for the short segment examined was calculated to be 2.19%.

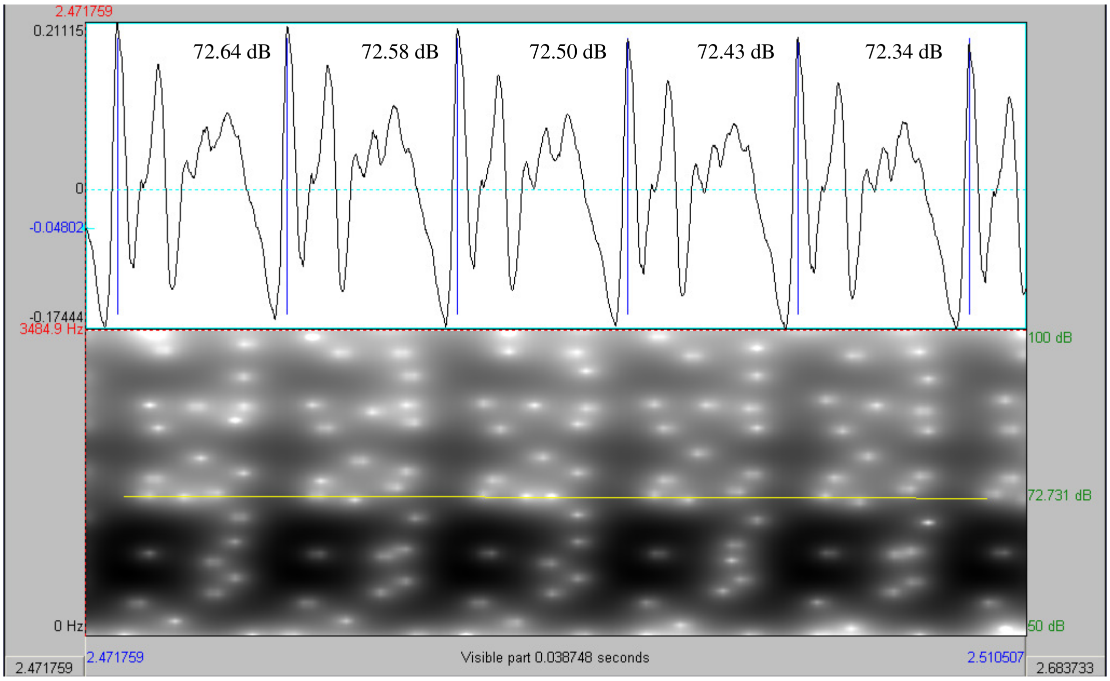

Figure 6. An illustration of a short segment extracted from a sustained phonation of the vowel /a/, produced by a male speaker. Vertical blue lines on the time-wave display represent start and end of individual phonatory cycles. Maximum intensity measurements for each cycle (in dB) are presented.

Shimmer is considered most appropriate for analysis of sustained phonations of vowels. However, when continuous speech is examined and voluntary changes in amplitude appear (i.e., prosody), the use of perturbation measures that include a smoothing factor is considered preferable (Baken, 1987). *Amplitude Perturbation Quotient (APQ)* is an acceptable measure for this purpose. It calculates a relative evaluation of the period-to-period variability of the amplitude within the analyzed signal, smoothing across 11 cycles. Specifically, it averages the differences between the amplitude values of a middle cycle to those of five preceding cycles and five successive cycles. Then, the same procedure is repeated for all series of 11 adjacent cycles, using a moving window. In the present study, APQ measurements were performed using the MDVP.

Further recommended reading:

Basic Concepts in acoustic analysis of voice:

Baken, R.J. (1987). *Clinical measurement of speech and voice*. Needham Heights, MA: Allyn and Bacon.

Examples of manuals for clinical and research oriented use of Praat:

1) http://www.fon.hum.uva.nl/praat/

2) http://www.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf

References

Baken, R.J. (1987). *Clinical measurement of speech and voice*. Needham Heights, MA: Allyn and Bacon.

Costello, J. M. & Ingham, R. (1984). Assessment strategies for stuttering. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions* (pp. 303-333). San Diego: College Hill Press.

Fletcher, H. & Munson, W.A. (1933). Loudness, its definition, measurement, and calculation. *Journal of the Acoustical Society of America, 5,* 82-108.

Howell, P., Au-Yeung, J. & Pilgrim, L. (1999). Utterance rate and linguistic properties as determinants of lexical dysfluencies in children who stutter. *Journal of the Acoustical Society of America, 105*, 481-490.

Most, T., Amir, O. & Tobin, Y. (2000). The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech, 43,* 295-308.

Stevens, S.S. & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology, 53,* 329-353.

Walker, J. F., Archibald, L. M. D., Cherniak, S. R. & Fish, V. G. (1992). Articulation rate in 3 and 5 year old children. *Journal of Speech and Hearing Research, 35,* 4-13.