Interaction between formant and harmonic peaks in vowel perception.

Hector Raul Javkin, Hynek Hermansky and Hisashi Wakita

Speech Technology Laboratory, Santa Barbara, California

ABSTRACT

The listener of a voiced vowel receives a signal consisting of formant-modulated harmonics. How this information is used in deriving both vowel timbre and resulting vowel identity is still not well understood. The suggestion by Klatt (1985,1986), that listeners perceive the actual resonance peaks, is contradicted by many works, including Mushnikov and Chistovich (1971) and Carlson, Granstrom and Fant (1975) who proposed weighted averages of neighboring harmonic peaks as the correlates of perceived vowel quality. Our perceptual experiments and re-analysis of the formant difference limen experiments of Flanagan (1955) and Nord and Sventelius (1979), support an interaction between formants and harmonic peaks in vowel perception.

INTRODUCTION

Although the influence of fundamental frequency on the perception of vowels is by now generally accepted [1,2,7,9, etc.], Klatt [3,5] has recently suggested that subjects respond to formant peaks without being affected by the location of the harmonic peaks determined by the fundamental, although he did find evidence for a normalization related to F0.

We found surprising support for the role of harmonic peaks in vowel perception in difference limen data shown in figures 1 [3], and 2 [8]. Both works provide the original measurement points along with the interpolated sensitivity curves. The measurements required a generous amount of interpolation to obtain smooth curves. If one takes into account the frequencies of the harmonic peaks in examining the published graphs, the origin of some of the output points can be hypothesized. The experiments were carried out with vowel formants from 500 to 1800 Hz, and F1 values ranging from 250 to 350 Hz in 25 Hz steps. Their results showed that their hypothesis worked the best among those examined, although its prediction is quite different from the perceptual data when F0 = 100 and F1 = 300, i.e., when the formant coincided with one of the harmonics.

A less compressed scale such as magnitude or intensity will increase the contribution of the strongest harmonic and can increase the correlation between the output of Carlson et al's equation and their experimental data. It should be noted, however, that an expanded scale is functionally similar to using the same scale but with the addition of some form of peak enhancement.
MATCHING EXPERIMENT

To test whether a different scale would yield results closer to those of human listeners, a matching experiment was conducted. Our aim here was to avoid the effects of categorization that occur in vowel perception and investigate the psychophysical effects. Accordingly, single-formant stimuli with a single resonance driven by a pulse train with a flat spectrum were synthesized. F0 was kept constant at 200 Hz. One set of stimuli had peaks ranging from 600 to 800 Hz in 20 Hz increments, the other set had the same increments, but ranging from 2000 to 2200 Hz. Both sets were prepared with three stimulus intervals: 200, 500, and 1000 Hz, for a total of 66 stimuli. Durations of the single-formant stimuli were 500 msec with 40 msec leading and 70 msec trailing edges, while the tones had a 500 msec duration and 60 msec leading and 120 msec trailing edges. The inter-stimulus interval was 200 msec.

The presentation of the stimuli and the recording of the responses were performed by a computer with a 10-bit digital-to-analog converter using a sample rate of 10 kHz with the output appropriately filtered. The stimuli were presented in different randomized orders to different subjects, who listened through earphones inside a sound-treated room. Subjects set the loudness of presentation to a comfortable level, and the level was checked visually after each subject completed the experiment. None of the subjects reported any hearing pathology. For each trial, subjects heard one of the single-formant stimuli followed by a pure tone. Their task was to match the timbre of the pure tone, adjusting the frequency of the sine wave, using keys on the computer keyboard. When they indicated satisfaction with each adjustment, their last adjusted value was automatically recorded in a computer file and the next trial began.

RESULTS OF MATCHING EXPERIMENT

Twelve subjects participated in the experiment. The task proved quite difficult for some subjects, and two of them were eliminated after complaining of the difficulty and giving up a third of the responses at the response limits. The results for the different bandwidths did not differ significantly but were noisy. The results were binned (within 20 Hz approximately two standard deviations) of the presented stimulus to limit somewhat the distorting effects of outliers. This meant that, for example, responses greater than 720 Hz to a 600 Hz stimulus were dropped from the data. Because of the band limitations in the presented stimuli and in the possible subject responses, points far away from the stimuli would severely distort the means. In addition, given a fundamental frequency of 200 Hz, a response of more than 750 Hz to a stimulus with a formant at 600 Hz might be the result of approaching the harmonic at 800 Hz. The results for the three bandwidths were combined in table 1, showing the results of the experiment for stimuli in the F1 range, and in table 2, showing the results for stimuli in the F2 range.

<table>
<thead>
<tr>
<th>Frequencies (Hz)</th>
<th>Mean (Hz)</th>
<th>Standard Deviation (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>F2</td>
<td>F3</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>------------------------</td>
</tr>
<tr>
<td>200</td>
<td>600</td>
<td>1000</td>
</tr>
<tr>
<td>600</td>
<td>1000</td>
<td>1500</td>
</tr>
</tbody>
</table>

**Table 1.**

**Table 2.**

**Figure 5.**

**Figure 6.**

The results for both sets are similar in their endpoints, although the data for the F2 range shows a less smooth pattern than the data for the F1 range. Both show a tendency for stimuli with harmonics close to the formant peak to attract responses and also for responses to slow a "plateau" when the formant is equidistant between harmonics. The responses would approximate a straight line if subjects were responding to the location of the formant peak without regard to the location of harmonics, so that the experiment confirms the effect of harmonic peaks. Nevertheless, the experiment does not confirm the predictions of Carlson et al.

CONCLUSIONS

It is clear from the results reported here, as well as from the majority of the experimental literature, that the location of harmonics plays a role in the perception of vowels and, more specifically, that harmonic peaks which collocated nearly coincident with formants tend to attract judgments of formant location. This effect appears to be too strong to be represented by a weighted average of the two most prominent harmonics in the bandwidths. Such an average can be improved by using a different scale, effectively expanding the differences in audiograms. Although the results reported here are still somewhat sketchy and must be considered with caution, they support the idea that such an expansion is necessary to describe the response of the human auditory system.

REFERENCES


ACKNOWLEDGEMENTS

We would like to thank Ted Applebaum, Jared Bernstein, Gregory De Haan, Brian Hanson, Dennis Klatt, Katia McClain, Lorio Mendes, Paul Neytrick, Ben Reeves and Kathy Smith for valuable discussions and help in various aspects of this paper.