Recording of synthetic vowels and their spectra

Rischel, J.

journal: STL-QPSR
volume: 6
number: 3
year: 1965
pages: 025-026

http://www.speech.kth.se/qpsr
II. SPEECH SYNTHESIS AND SPEECH PERCEPTION

A. RECORDING OF SYNTHETIC VOWELS AND THEIR SPECTRA*

J. Rischel**

At the Speech Transmission Laboratory a series of vowels has been synthesized and subjected to spectral analysis.

The synthesis was performed on OVE II. The formant frequencies were set manually and checked by means of a Hewlett-Packard tone generator and an electronic frequency counter. Each vowel was recorded on an Ampex tape recorder after an analysis had been made by means of a filter bank spectrum analyzer.

The spectrum analyzer, which has been described by J. Liljencrantz in STL-QTSR 2/1965 (1), had not previously been used for large-scale processing of phonetic material, so that it is of some interest to study its presentation of various types of vowel spectra. A high frequency resolution of the spectrum up to and including the fourth formant was considered desirable, and the so-called group analysis method was chosen. The vowel signal from the synthesizer was applied to a bank of 40 band-pass filters which together covered a frequency range of 500 c/s. By means of a heterodyne process this filter bank was made to cover different parts of the pertinent spectrum (0-500 c/s, 500-1000 c/s, etc.), the analyzed frequency range extending from 0 to 4 kc/s. With $F_0$ fixed at 100 c/s a well-defined formant pattern was obtained, although the remarkable feature of the curve is an excellent resolution of the fine structure of the spectrum because of the very close spacing of the analyzing filters.

The dynamic range of the analyzer is specified by J. Liljencrantz as being 60 dB quantized in steps of 0.5 dB. This certainly represents a considerable improvement as compared to current equipment such as the Sonagraph. With the 60 dB range it is possible to record the formant pattern of probably all vowels without pre-emphasis, although in the case of narrow back vowels the energy above $F_2$ drops to the level of background noise.

* The author is indebted to Mr. P-O. Birnon and Mr. Á. Florén of the Speech Transmission Laboratory for technical assistance in the recording of the material referred to here.

** Institut for lingvistik og fonetik, København, Denmark.
Fig. II-A-1. Vowel of [o] type. Formants located at 350, 800, 2500, and 3500 c/s.
Fig. II-A-2. "Danish" [æ]. Formants located at 550, 1950, 2500, and 3500 c/s.

Fig. II-A-3. Vowel of [ε] type. Formants located at 650, 1300, 2500, and 3500 c/s.
from which only the most intense partials stand out, cp. Fig.
II-A-1 exemplifying a vowel of [o] type. In vowels with a more
even energy distribution also the valleys between the formants

One purpose of performing the analysis described here
was to get a more comprehensive stock of formant envelopes than
those generally available. Sixty different vowel spectra were
recorded and it is assumed that these may be useful in providing
evidence concerning the discrepancies between pole frequencies
and estimated peaks of spectral energy. Since the exact pole
frequencies are known for each of the synthesized vowels, informa-
tion is obtained concerning the envelope shapes of a variety
of formant configurations.

The vowels recorded represent formant configurations
typical of close, half-close, half-open, and open front vowels
with and without lip-rounding and back vowels with lip-rounding
plus some central vowels. In addition to this "standard" set of
vowels some series were produced with fixed higher formants and
with a stepwise variation of $F_1$. These vowels are (as defined by
the frequency location of $F_2$ and $F_3$):

```
<table>
<thead>
<tr>
<th>$F_2$</th>
<th>$F_3$</th>
<th>$F_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i] 2100 c/s</td>
<td>3000 c/s</td>
<td>3500 c/s</td>
</tr>
<tr>
<td>[e] 2200 c/s</td>
<td>2750 c/s</td>
<td>3500 c/s</td>
</tr>
<tr>
<td>[y] 1850 c/s</td>
<td>2100 c/s</td>
<td>3500 c/s</td>
</tr>
<tr>
<td>[ʊ] 1700 c/s</td>
<td>2250 c/s</td>
<td>3500 c/s</td>
</tr>
<tr>
<td>[ɔ] 1500 c/s</td>
<td>2500 c/s</td>
<td>3500 c/s</td>
</tr>
<tr>
<td>[o] 800 c/s</td>
<td>2500 c/s</td>
<td>3500 c/s</td>
</tr>
</tbody>
</table>
```

In the case of [i], [e], $F_1$ was varied in steps of
25 c/s from 200 to 450 c/s for [i] and from 250 to 500 c/s for
[e]. For the other vowels a variation in steps of 50 c/s was
chosen (except that fewer samples were recorded of the central
vowel). In each case $F_1$ was varied within the limits consistent
with the vowel retaining a reasonably natural vowel-like quality.
These vowel series are intended primarily for use in connection
with research on vowel perception.

Reference:

(1) Liljanorants, J.: "A filter bank speech-spectrum analyzer",