Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

OVE II simulation of errors in formant tracking
Campanella, S. J. and Coulter, D. C.

journal: STL-QPSR
volume: 3
number: 3
year: 1962
pages: 028-029

http://www.speech.kth.se/qpsr
B. OVE II SIMULATION OF ERRORS IN FORMANT TRACKING

This is a technical note on an experiment performed by S.J. Campanella and D.C. Coulter during their visit to R.I.T. from 3 September to 11 September, 1962.

It was desired to ascertain by re-synthesis on OVE II the effect of various corrections of formant-tracked compressed speech signals. Specifically, it was desired to find out which of the parameters $F_1$, $F_2$, $F_3$ were contributing the greatest error in the perceived sound. The utterance chosen for this experiment was the phoneme pair /ga/, which, when processed through the Molpar system was most often perceived as /da/ in forced choice decisions between /b/, /d/, and /g/.

As shown in the attached spectrographs (Fig. II-4) for the original and formant tracked utterance for /ga/, three principal errors were noted. First, the entire $F_1$ trace was about 100 cps low in its position. Secondly, $F_2$ possessed a "hook" distortion during the initial 16 to 24 milliseconds, thus lowering its initial frequency position, and thirdly $F_3$ is erroneous and even possessed a reverse initial transition.

The experiment was initiated by deriving the formant trace from spectral analysis and then replotting them on the OVE II calibrated function chart in conducting ink. The examples plotted consists of the original utterance, the processed version, and a third version which was the processed speech to be subjected to various corrections. The objective was to allow the corrected version to be compared with both the original and the normally processed speech. During the experiment, aspirative and consonantal cues were not used, although it is known that these may contribute to the perception of the /g/ sound. This was done purposely, since it was desired to evaluate the effect of voiced cue degradation on stop consonant perception.

The initial synthesis produced speech that was acceptably close to the source samples except that the aural quality was that of OVE II for both the original and processed speech, and the absence of unvoiced cues made the original /ga/
FREQUENCY - ORIGINAL SPEECH

FREQUENCY - SYNTHETIC SPEECH

ORIGINAL AND SYNTHETIC SPEECH MICROGRAMS
OF /ga/. SYNTHETIC SPEECH PERCEIVED AS /da/

Fig. II-4.
sound a little less definite. A third sample was now constructed which was identical to the second (processed speech) except that formant three was corrected to give the proper upward /g/ transition. The effect was to reduce the /d/ impression, but not to substantially increase the impression of the /g/ sound. Rather, the effect was more towards the liquid /y/ sound, yet probably possessing under a forced choice /b, d, g/ something closer to the /g/. A second correction tried was that of delaying the onset of amplitude until the "hook" in the F₂ trace, produced by the rise time associated with the F₂ tracker and low-pass filter, was not voiced. It was thought that perhaps delaying the amplitude onset to allow time for the trackers to get into position might permit acceptable articulation for initial stop consonant such as the subject sample. The result of this experiment was not decisive, i.e. further loosened the /d/ impression without strengthening that of the /g/. The third experiment consisted of substituting the F₁ trace from the original speech for the synthetic one. The audible effect was two-fold. First, a relatively good /ga/ was produced, and second, an improvement in the vowel color of /a/ was obtained. As a final test, the point of initiation of A₀ excitation was again advanced to include the erroneous initial hook of the second formant in the voiced event. This change produced surprisingly little alteration in the acoustical output, and one could still perceive a rather good /gc/. This leads one to conclude that, for this example at least, the information for manner of articulation contained in the F₁ transition is extremely significant in creating the perception of the perception of the consonant /g/ in the /ga/ utterance.

In conclusion, it appears that to produce acceptable stop consonants based on voiced transitions requires excellent accuracy and fidelity in tracking the values of the first three formants.

References:


S.J. Campanella, D.C. Coulter