Experiments with nasal synthesis

Nord, L.

journal: STL-QPSR
volume: 17
number: 2-3
year: 1976
pages: 014-019

http://www.speech.kth.se/qpsr
B. EXPERIMENTS WITH NASAL SYNTHESIS

L. Nord

Abstract

The report summarizes some results from analysis and synthesis work with nasals. Nasal consonants were synthesized, using a modified, simulated version of OVE III speech synthesizer that included an extra pole-zero pair in series with the vowel branch. A slight improvement in quality of the nasals was obtained. Some general results from synthesis of nasals, using different strategies are reported. The results lend support to the present strategy used at the Dept. of Speech Communication.

Introduction

The following report summarizes results from analysis and synthesis work with nasals. The study had two purposes. The first question I tried to answer was whether the spectral shape of the synthetic nasal murmur recommended to be used for OVE III speech synthesizer is a too coarse approximation compared with the complex shape of nasal murmur in real speech. It was thus tested whether a modification of the circuit configuration of OVE III by the introduction of an extra pole-zero pair in series with the vowel branch would improve the quality of the nasal sounds.

Furthermore, different strategies for synthesizing nasals were compared in a fairly systematic way. The aim was to ascertain the demands of complexity for the synthesis patterns needed, when producing nasals in various contexts.

Background

The complex shapes of the spectral envelopes of nasal consonants are well known, see Fant (1960); Flanagan (1965); Fujimura & Lindqvist (1971); Lindqvist-Gauffin & Sundberg (1976). The mouth and the sinuses of the nose form shunting cavities to the main path, namely the pharynx and the nose tubes, thereby adding pole-zero pairs to the frequency spectrum of the sound produced. A typical pre-emphasized, cepstrally smoothed amplitude spectrum of the nasal /m/ is shown in Fig. II-B-1. Assumed locations of the lower poles and zeroes are indicated by "x" and "o", respectively. Note the broad low-frequency peak, probably containing two poles and a zero, the marked spectral zero below 1 kHz due to the mouth cavity shunt and the two resonances around 1 kHz, all contributing
Fig. II-B-1. FFT-spectrum of /m/. Pre-emphasized and with cepstral smoothing. Assumed locations of the lower poles (x) and zeroes (o) are marked.
to form a smooth spectrum very different from a vowel spectrum with its marked formant peaks and little energy in-between.

An analysis made with several speakers shows a large variation of the spectral shape of the nasal murmur which would indicate that the finer structure is of little relevance to the phonemic information. Nevertheless, perceptual experiments (Nord, pp. 5-8 in this issue of STL-QPSR) where nasals have been interchanged show that listeners can use information from the nasal segment in order to identify the nasal (i.e. the nasal murmur itself and not only the formant transitions of the surrounding vowels contributes to the place information).

Analysis. Matching procedures

The synthesis has been supported by a close analysis of one subject's nasal production with complimentary comparisons with other speakers. The acoustical analysis consists of broad-band spectrograms and Fast Fourier Transform analysis (FFT) on a computer (Liljencrants, 1969). A very good dynamic picture of the acoustic event is obtained by plotting consecutive FFT sections, see Fig. II-B-2. This facilitates a tracing of the formant movements also in low energy portions of the signal. Note the continuity of F1 from the vowel into the nasal and the low-frequency "nasal resonance" already present in the vowel preceding the nasal.

To get a more complete picture of the pole-zero locations a spectral matching was performed using a computer program where the desired number of poles and zeroes could be specified.* Two types of restrictions were imposed on the matching procedures to be able to use the results directly for the synthesis: (1) The number of singularities were restricted during the matchings of the different spectra. (2) The poles and zeroes of the matching spectrum corresponding to the correction for higher poles were held fixed. In some cases then, an adding of more pole-zero pairs would give a better match, especially at higher frequencies.

Fig. II-B-3a shows a matching of a schwa vowel produced by the OVE III synthesizer** and Fig. II-B-3b shows the matching of a spectral sample of a natural nasal consonant where the inventory of matching

* POLAR, a computer program written by J. Liljencrants
** a simulated version of OVE III
Fig. II-B-2. Consecutive FFT-sections of /ana/. 10 msec between each section. A tracing of the formant movements is indicated.
Fig. II-B-3a. Spectral matching of a synthetic schwa vowel. Singularities are marked with frequency and half the bandwidth, poles upwards and zeroes downwards. 4, 5, 6, 7, 8 constitute the correction for higher poles, including F4 (7 and 8 are real zeroes).

Fig. II-B-3b. Spectral matching of /n/ in an /a-a/ frame.
singularities has been increased with three pole-zero pairs.

In Fig. II-B-4 a comparison is made with some results obtained by Fujimura (1962) and Fant (1960). As can be seen the overall similarities are good in spite of the restrictions mentioned above. The difference in articulation between an American alveolar /n/ and a Swedish dental /n/ is probably reflected in the different positions of the zero due to the mouth cavity shunt (1.6 kHz for the American, 1.0 kHz for the Swedish).

Experiment 1

Synthesis

The complex shape of the nasal spectrum found in real speech raises the question whether a more detailed synthetic match would improve the quality of the nasal sounds.

The OVE III configuration is based on an all-pole model of the vocal tract (Liljencrants, 1968). An extra nasal resonance (N1) is also used simultaneously with the vowel formants for nasals. This nasal branch is connected in parallel with the vowel branch in a way that results in a spectral dip, a zero, in the vicinity of N1 and F1. The configuration thus includes three formants controlled in frequency and bandwidth and a pole-zero pair used for emphasizing the low-frequency band.

A characteristic feature of the matched spectra, see Fig. II-B-4, is the pole-zero pair resulting from the mouth cavity shunt, giving an emphasis to the envelope at about 1 kHz for /m/ and somewhat higher for /n/.

Specifically we therefore wanted to know if a modified synthesizer including an extra pole-zero pair in series with the vowel branch would improve the nasal quality.

A simulated version of OVE III, SIMOVE*, was used for the synthesis experiments. This simplified the changes of the resonance configuration like an adding of a pole-zero pair into the changing of a few figures in the computer program. A disadvantage, however, was the difference in quality between the software and the hardware synthesis. The simulated synthesizer had a softer quality, more close to a nasal voice which made comparisons between the two impossible. Therefore all comparisons between different stimuli were made with SIMOVE.

* The simulated version of OVE III is written by J. Liljencrants
Fig. II-B-4. A comparison of modelled and matched nasal consonants.
Synthetic stimuli with and without the extra pole-zero pair were made from analysis and matching data. The nasals /m, n, η/ were placed in symmetrical vowel frames /a-a, I-I, U-U/. An example of the stimuli tried with the extended SIMOVE is shown in Fig. II-B-5.

Results

Listening tests show that subjects are able to distinguish the stimuli with and without a pole-zero pair and to a varying degree prefer the more complex pattern. However, the improvements are in most cases so small that the outcome of these tests does not justify a change in OVE III configuration.

Experiment 2

Different synthesis strategies

The second intention of the study was to try different synthesis strategies in a systematic way to find out the relation between synthesis pattern complexity and nasal quality. The underlying idea being that nasals in different contexts would require synthesis patterns of various complexity.

Earlier attempts

A few earlier experiments describing synthesis techniques for nasals will be discussed. Carlson, Granström, and Pauli (1972) made an experiment there, they synthesized nasals using OVE III with an extra resonance in the nasal branch. NV-syllables were made with the nasal containing one or two resonances (with a zero in-between) only. The results show that in some nasal-vowel combinations listeners prefer the more complex nasal containing two resonances.

The problem of how to synthesize sounds containing spectral zeroes with a series synthesizer are dealt with in two articles: Mártony (1964) tried a strategy of only using the vowel formant branch for making nasals, with OVE II. The nasals and the nasalized vowels of NV-syllables were spectrally matched by frequency and bandwidth modifications. See also a similar study by Nakata (1959).

The present strategy used in the rule-synthesis program made by Carlsson and Granström at the Dept. of Speech Communication involves the use of both the vowel and the nasal branch for the nasals. Changes in
Fig. II-B-5. Synthesis pattern and spectrogram of /ama/ containing formants, nasal resonance (N1), and an extra pole-zero pair (x, o). AV is the amplitude of the vowel branch, AN the amplitude of the nasal branch.
bandwidths are used to meet the effect of pole-zero pairs altering the
spectral amplitude relations.

**Synthesis**

Fig. II-B-6 shows examples of different strategies tried in VNV
stimuli. The complexity increases from top to bottom of the diagram,
starting with only one resonance (N1) used in the nasal gradually down to
the complex form already described and used in the first part of the ar-
ticle, with vowel formants, a nasal resonance and an extra pole-zero
pair (x, o). The nasalization of the surrounding vowels are also syn-
thetized with the nasal resonance and the pole-zero pair. (For the vowel
/a/ the pole is placed at about 2 kHz and the zero at 2.35 kHz, thus re-
ducing the level of F3.)

The different synthesis strategies were tested on /ama, ana/ and /Inl/.
To facilitate a comparison between the stimuli of which some differed
very little in quality, a tape was prepared with the stimuli occurring in
different combinations with each other.

The material was presented to trained phoneticians at the Dept. of
Speech Communication and evaluated at informal listening sessions. As-
pects of intelligibility as well as naturalness were considered.

**Results**

The results are mostly of qualitative nature and give some indica-
tions for the specification of synthetic nasal patterns. Only a few remarks
will be made here.

A general tendency is that the more complex patterns for the nasal
consonants, i.e. patterns that more faithfully reproduce the original
speech wave, always are preferred in varying degrees. It is also the
case that the complexity needed for the acceptance of the nasals is very
different for the different test stimuli. Thus the word /ama/ is accepted
by many listeners already when it is synthesized with the vowel formants
only. /ana/ will easily be confused with /ala/ when made only with for-
mants; an added nasal resonance in the low-frequency band is needed to
perceive a nasal quality. The same applies for /Inl/ although most list-
eners don’t accept the /Inl/ stimulus until also the extra pole-zero pair
is added. The greater tolerance for variation of the /ama/ specifications
is probably partly explained by the fact that the phoneme inventory lacks
a labial correspondent to the dental /l/.
IN THE NASAL:

- nasal resonance formants
- nasal resonance
- extra pole-zero pair

Fig. II-B-6. Schematic synthesis patterns for /ama/. Examples of synthesis strategies.
In general those stimuli that have a continuity at low frequency by a nasal resonance going all through the stimulus combined with a reduced F1 level by a bandwidth increase at the nasal-vowel boundary are preferred. The results thus lend support to the present synthesis strategy used.

Conclusions

Nasal consonants synthesized with a modified (simulated) OVE III, including an extra pole-zero pair in series with the vowel branch with the aim of better matching the nasal murmur are found to be improved in quality. The marginal improvements compared to the present strategies used, however, do not justify a change of the OVE III configuration.

The comparisons between different synthesis strategies for nasals give some detailed information concerning the relations between pattern complexity and perceived nasal quality. An evaluation of different strategies lends support to the strategy presently used where the vowel and nasal branch are used simultaneously and the bandwidth of the first and the second formants are increased in order to better match the nasal murmur and the nasalized vowels.

References


