Umeå University, Department of Philosophy and Linguistics
PHONUM 9 (2003), *145-148*
Available online at http://www.ling.umu.se/fonetik2003/

Turning wine into water: Can ordinary speech be artificially nasalized?

Erik Eriksson[1], **Jan van Doorn**[2,3], **Kirk P. H. Sullivan**[1,4]

[1]Department of Philosophy and Linguistics, Umeå University, [2]School of Communication Sciences and Disorders, The University of Sydney, [3]Department of Clinical Sciences, Umeå University, [4]Department of Philosophy and Linguistics, Umeå University

Synthetic speech has recently been used to study resonance and voice disorders. The advantage of synthetic speech is that it becomes possible to control and artificially manipulate acoustic variables related to a specific perceptual feature without confounding effects from other co-occurring problems often found in disordered speech. To date, synthesis has been restricted to vowels in isolation or in single words. An analysis-synthesis technique has been used that starts by making a spectral match to actual disordered speech. The study described here has investigated the feasibility of using another technique that may be suitable for connected speech contexts. The technique involves spectral manipulation of vowels within samples of connected natural speech. Several variations were tested to mimic the nasal cavity's effect on the speech signal: direct manipulation of spectral features extracted from linear predictive coding analysis, filtering with a pole-zero filter and adding a pole-zero pair to the transfer function. All spectral modifications were applied to vowels within sentences of natural speech. Problems that were encountered with the techniques are discussed. Further a database of pilot sentences has been prepared for perceptual evaluation of their resonance quality and comparison of the different types of modification.

1. Introduction

Hypernasality is often encountered in speech pathology, as evidenced by the large body of literature pertaining to this subject. It is a feature of cleft palate speech, dysarthria, and deaf speech. Speech is deemed to be hypernasal (excessively nasal) when oral-nasal coupling within the vocal tract is increased sufficiently to result in a perceptually significant change in the speech sound (Curtis, 1968).

Physiological coupling of the oral and nasal cavity alters the filtering properties of the vocal tract, producing changes in the speech spectrum. Although there is a reasonably consistent set of acoustic features associated with hypernasal speech (Kent, Liss & Philips, 1989), no one set of features has yet been identified that can be applied to all vowel types. In general, the main features of nasalization are changes in the low frequency regions for the speech spectrum, where there is very low frequency peak with wide bandwidth along with the presence of a pole-zero pair due to acoustic coupling (Hawkins & Stevens, 1985)

There are numerous reports on the difficulty of making reliable perceptual judgments on the degree of hypernasality (eg Counihan & Cullinan, 1970). While hypernasality is considered to be essentially a vowel based phenomenon (Schwartz, 1979), its perception is

often confounded by other related speech features such as poorly articulated pressure consonants. Perceptual rating of hypernasality is inherently difficult, and it is made even more difficult because it occurs typically in conjunction with abnormalities in pitch, loudness, voice quality, and/or articulation (Spriestersbach, 1955). These coexisting features affect the perception of hypernasality (Moll, 1968). Thus it is difficult to investigate the effects of hypernasality in isolation. For instance, it has been found that hypernasality affects speech intelligibility (Seikel, Wilcox & Davis, 1990). However, it is difficult to evaluate whether hypernasality on vowels alone is responsible for reduced intelligibility, or whether other co-existing speech problems also influence the intelligibility.

Several recent studies have addressed this problem by using synthetic nasal vowels to study nasal speech (Hawkins & Stevens, 1985; Huffman, 1990; Chen, 1995; Zraik & Liss, 2000). Using synthetic nasal vowels removes interference from other sources and allows single factor investigations on various aspects of hypernasality to be conducted. In the studies to date, synthetic nasal vowels (in isolation or in single word contexts) have been produced using analysis-synthesis techniques that start by making a spectral match to actual disordered speech.

An alternate method of generating artificial hypernasal speech would be to alter the acoustic features of the vowels within natural connected speech so that they represent nasalized vowels. The aim of the investigation reported in this paper was to test whether perceptually nasalized speech could be produced by modifying the spectral features of the vowels to resemble nasalized vowels.

2. Method

Ten sentences for 12 different speakers were modified using three different techniques. Each sound file was altered several times for each kind of alteration (formant shift, pole-zero-pair filtering and pole-zero-pair adding), within a range of values, generating different outputs every time.

2.1. Original speech waves

The original natural speech recordings were 10 selected sentences from the Australian National Database of Spoken Language (ANDOSL). Recordings from six male and six female native speakers of Australian English were used. The sentences selected contained either no nasal consonants (for 4 sentences) or one nasal consonant (6 sentences).

2.2. Speech modifications

Several types of modification were tested to mimic the nasal cavity's effect on the speech signal: direct manipulation of spectral features extracted from linear predictive coding analysis, filtering with a pole-zero filter and adding a pole-zero pair to the transfer function.

2.2.1. Preparation of speech samples.

The speech samples were normalized, down sampled to 10 kHz and bandpass filtered from 70 Hz to 5 kHz. Voicing detection was performed on the speech samples, based on a 10 millisecond window and all consecutive voiced and unvoiced segments compressed into one (silence was coded as unvoiced speech). A linear prediction (LP) analysis of the order of 14 was performed on all voiced segments with a window of 20 milliseconds and a 50 % overlap.

2.2.2. Modification of speech samples.
The alterations applied to the original sound files were constructed to mimic nasal coupling. Each speech sample's voiced segments were altered using the following methods:
(1) Modifications of the first vowel formant (F1) using LP parameters.
The first formant (F1) of the LP spectrum envelope was matched to the LP coefficient and changes applied to that coefficient. The altered coefficient was then re-entered into the LP folder, replacing the old one. If the formant tracking failed to find a peak (pole), or if the pole bandwidth was too wide the segment was coded as unvoiced. The F1 modification used eight different changes to the coefficient associated with F1.
(2) Pole-zero filtering.
A pole-zero filter was constructed based on frequencies and bandwidths found in the literature (Hawkins & Stevens, 1985; Kingston & MacMillan, 1995; Zraick & Liss, 2000). The voiced segments were then filtered using one of several pole-zero pairs. The pole-zero filtering technique made 20 different modifications by changing the pole-zero pair over a range of frequencies.
(3) Pole-zero pair adding.
Similar pole-zero pairs as constructed in (2) were added to the LP folder, instead of filtering the voiced segments. Unstable filters (after adding of pair) were stabilized. The addition of a pole-zero pair to the transfer function of the original speech tested two different pairs.
2.2.3. Resynthesis of modified speech signals.
The final stage of the alteration process involved resynthesis of the speech signal. In methods (1) and (3) the newly constructed LP coefficients filtered the residual signal from the original LP analysis. In method (2) filtered (voiced) and unfiltered (unvoiced) segments were concatenated together. A final instability check was performed for each segment and if the segment was unstable, it was deleted. Additionally, the impact of the alterations on the quality of the speech signal was inspected informally during testing of the methods.

3. Results and Discussion

Prior to an investigation of the success of the methods in achieving nasal speech quality, it was necessary to check the quality of the resynthesized speech samples. During checking it was found that the output signal was cluttered with noise and chirps, resulting in poor quality especially for the LPC alteration and pole-zero pair adding methods. Steps where taken to reduce the noise but the results were still unsatisfactory. There are a number of likely reasons for this, with the major one being the voicing decision. The LPC analysis is preferably implemented in a pitch synchronous environment with the unvoiced frames unaltered. However, as the voicing decision is not exact, the frames might be incorrectly marked as voiced, even though they are unvoiced. Further, the voiced-unvoiced boundary may have been erroneously calculated, resulting in frames with both voiced and unvoiced signal inside.
The noise could also have been a result of too large shifts in the formants. That is, the movement of a single pole causes the transfer function to become almost unstable (i.e. poles very close to the unit circle). These were not detected with a stability check as they lay within the unit circle but could still affect the signal unfavourably. This was shown during testing by an increase in noise as the frequency and bandwidth shifts increased.

Another reason for the quality reduction could be within the reconstruction of the signal. The signal was divided into frames before alteration and was then reconstructed by concatenation. This could cause noise, as the difference between the last signal point in the first frame and the first signal point in the next frame might have increased. This was also true when a frame was removed due to instability. There are ways to solve this issue and

Chappell & Hansen (1998) have presented means to smooth the transitions.

This feasibility study has indicated that modification of vowels within natural speech can potentially produce artificially nasalized connected speech. Problems with quality of the reconstructed speech need to be further investigated so that the next stage of perceptual testing for validity of the nasality modifications can be conducted.

4. Acknowledgments

Sincere thanks to Wendy Hooper for her thorough literature search prior to the start of this project.

5. References

Chappell, D. T. & Hansen, J. H. L. (1998) Spectral smoothing for concatenative speech synthesis. In *Proceedings of 1998 International Conference on Spoken Language Processing*, 5, 1935-1938.

Chen, M. (1995) Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers, *Journal of the Acoustical Society of America,* 98, 2443-2453.

Counihan, D. T. & Cullinan, W. L. (1970) Reliability and dispersion of nasality ratings. *Cleft Palate Journal*, 7, 261-270.

Curtis, J. F. (1968) Acoustics of speech production and nasalization. In *Cleft palate and communication* (D. C. Spriesterbach & D. Sherman, editors), pp. 326-330. New York: Academic Press.

Hawkins, S. & Stevens, K. N. (1985) Acoustic and perceptual correlates of the non-nasal nasal distinction for vowels, *Journal of the Acoustical Society of America,* 77, 1560-1575.

Huffman, M. (1990) The role of F1 amplitude in producing nasal percepts, *Journal of the Acoustic Society of America*, 88 (Suppl 1), S54.

Kent, R. D., Liss, J. M. & Philips, B. J. (1989) Acoustic analysis of velopharyngeal dysfunction in speech. In *Communicative disorders related to cleft lip and palate. 3rd edition* (K. R. Bzoch, editor), pp. 258-270. College-Hill.

Kingston, J. & MacMillan, N. A. (1995) Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants: A detection - theoretic application of the garner paradigm, *Journal of the Acoustical Society of America*, 97, 1261–1284.

Moll, K. L. (1968) Speech characteristics of individuals with cleft lip and palate. In *Cleft palate and communication* (D. C. Spriestersbach & D. Sherman, editors). New York: Academic Press.

Schwartz, M. F. (1979) Acoustic measures of nasalization and nasality. In *Communication disorders related to cleft lip and palate* (K. R. Bzoch, editor), pp. 263-268. Boston MA: Little Brown.

Seikel, J. A., Wilcox, K. A. & Davis, P. J. (1990) Dysarthria of motor neurone disease: Clinicians judgements of severity*, Journal of Communication Disorders*, 23, 417-431.

Spriestersbach, D. C. (1955) Assessing nasal quality in cleft palate speech of children, *Journal of Speech and Hearing Disorders*, 20, 266-270.

Zraick, R. I. & Liss, J. M. (2000) A comparison of equal-appearing interval scaling and direct magnitude estimation of nasal voice quality, *Journal of Speech Language and Hearing Research,* 43, 979-988.