

# Nasality in musical sounds – it is not a frequency band

Robert Mores, Insa Malhotra

University of Applied Sciences Hamburg (HAW), 20099 Hamburg, E-Mail: [mores@mt.haw-hamburg.de](mailto:mores@mt.haw-hamburg.de)

## Introduction

A recent meeting of reputed violin researchers and luthiers in Cambridge led to this review. It seems that “nasality” is one of the frequently used but not well understood terms. Perceptual studies on nasality in sounds often conclude without significant results and there is no general model yet.

The violin research community still trusts the early definition of Dönnwald, who did a tremendous work in measuring more than 1000 violins and defining four characteristic energy bands for violins. One of these bands was assigned the nasal band, and in the latest publication the band ranges from 700 Hz to 1600 Hz [HEI03]. In the strings community, these bands serve as reference today as well as the assigned terminology. However, the speech processing community has established other acoustical properties (APs) to capture nasality, and clinical research has also established its own perspective on nasality. This paper summarizes findings from these fields as well as from some own studies.

## Nasality in Speech Processing

After a period of fragmented research the community settled with some well accepted APs. Early APs investigated are: (i) some prominence at 1kHz, an additional dip in the range between 700-1800Hz and a reduced A1, the amplitude of the F1, the 1<sup>st</sup> formant [Hou56], (ii) a resonance at 250Hz and a zero at 500Hz [Hat58], (iii) increase of bandwidth of F1, F1BW, and an extra formant at 2 kHz [Fan60], (iv) increased bandwidth for F1 but also for F2, changed amplitudes and frequency of F1, F2 and F3 [Dic62], (v) frequency-shift of F1, extra zero-pole around F1 [Fuj71], (vi) flattening of the spectra in the range of 300 to 2500 Hz [Mae82], (vii) prominence of an extra pole around F1 [Haw85], (viii) widening of the F2-F3 region plus two extra pole-zero pairs between 220 - 2150 Hz [Bog86]. In summary, early research results do not support a general model.

Today's most well accepted APs for nasality in speech are: (i) the standard deviation around center of mass in the band below 1kHz, and the percentage of time of observed extra poles at low frequencies [Glas85], (ii)  $df\_P0$  and  $df\_P1$ , that is the frequency of the nasal extra poles  $P0$  and  $P1$  with respect to the frequency of  $F1$  [Mae93], (iii)  $dA\_P0$  and  $dA\_P1$ , the amplitudes of the extra poles with respect to the amplitude of  $F1$  [Che95], (iv)  $F1$  bandwidth and other  $F1$  profile criteria, and the number of peaks above a threshold 40dB below signal peak, and two criteria relating the amplitude of the first formant to the first harmonic [Pru07]. Different sets of some 10 to 20 of these and other APs are usually taken as a knowledge-based parameter set to solve binary nasality classification tasks. Most of these studies deliver an accuracy between 60% and 90%. Some of the APs introduced by Glass in 1985 are now expressed by the nasal poles  $P0$  and  $P1$  around  $F1$ , and their relation to  $F1$  in terms of frequency and amplitude [Gla85].  $P0$  and  $P1$  are usually dominated by  $F1$  and  $F2$  and are difficult to separate, as shown in Figure 1. Even more difficult is the extraction of their bandwidth or amplitude. Pruthi has demonstrated classification results with an accuracy of up to 96 %, 78 %

and 70 % on the StoryDB, TIMIT and WS96/97 data sets, respectively, with an RBF kernel SVM [Pru07]. He also changed the paradigm of static sinus resonance frequencies and identified the interdependence of nasal APs and vowel quality.

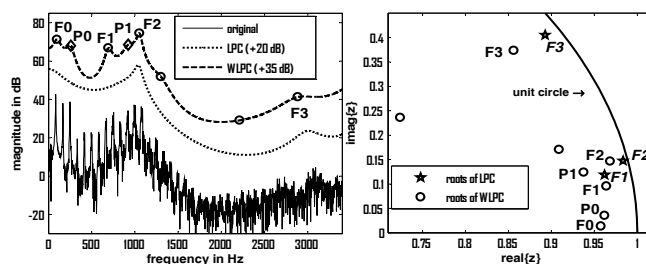


Figure 1: nasal speech signal ‘a’, left: frequency domain and its related LPC and warped LPC spectrum, right: roots of the coefficients in the z-plane

The speech community seems to have settled with the search for appropriate APs, most of which are located around  $F1$  and well below 1 kHz. There is full awareness that there are many more APs at higher frequencies, however, with little chance for modelling. Pruthi has shown in his simulations, that velum motion causes extra poles and zeros across the full range between 1 kHz and 3 kHz, depending on the size of the coupling area between the vocal and the nasal tract, and depending on the vowel context [Pru07]. This confirms the complexity issue and explains the problem of generalisation for APs at the higher frequencies. Today the speech community investigates additional cues to improve speech-to-text categorical tasks: phonetic context and murmur thresholds, and energy over time fluctuations [Ber07] [Haj04].

## Perceptual issues

Early perceptual studies deliver a likewise heterogeneous scenario: (i) A1 reductions [Hou56], (ii) individual poles and zeros in 1958 [Hat58], (iii) spectral flattening at low frequencies [Mae82], (iv) pole-zero pair insertion [Haw85], and (v) formant shifts [Bog86] have been presented to listeners to rate nasality.

Cross-language studies confirm the necessity of careful test design for perceptual studies: (i) for French and American English speaking listeners perceived nasality depends on vowel duration [Del68], (ii) perceived nasality is less obvious for syllables with a plosive environment than for syllables with a fricative environment [Lin61], (iii) perception of oral-nasal vowel distinction is categorical for Hindi speakers, and more continuous for speakers of American English [Bed82], (iv) British English speaking listeners prefer some murmur along with brief nasalization in the vowel, French speaking listeners prefer a longer duration of nasalization [Ste87], (v) nasal vowels presented in isolation or in oral context are more often correctly judged as nasal, than when presented in the original nasal context [Kra91]. In summary, language background will strongly bias test responses.

## Clinical studies

Yet another perspective on nasality opens when reviewing clinical studies. Whereas speech processing aims at speech or speaker-specific feature extraction, clinical research aims at diagnosis and therapy of speech problems or disabilities. This different focus and context has led to other approaches in terms of analysis, modelling and data bases. Clearly, nasality measurements are likely to fail for patients where the phonetic context is shifted due to conjoined cleft palates. Baken et. al. identified following APs for nasality: larger F1 bandwidth, frequency shifts of formants, an extra pole between 250 Hz and 500 Hz, an extra zero around 500 Hz, irregular extra poles between formants, and a lower total signal energy [Bak02]. Only some of these APs are similar to those found by the speech community. Zecevic's classification study with SVM uses the first four formants and their frequency, amplitude and bandwidth, ignoring the extra poles P0 and P1. The investigated data corpus NASAL contains more than 3000 sounds from 116 male, female and infant speakers, and the overall classification accuracy is well comparable with results in speech processing research [Zec02].

## Own studies between disciplines

In a brief study we extended the extraction method. F1, P0 and P1 are extracted using the warped LPC and a root solver on the LPC coefficients. Figure 1 demonstrates the superiority of warped LPC against LPC when searching for properties on the low frequency side. In the z-plane, bandwidth and frequency of P0 and P1 can be measured even when masked by F1, see Figure 2. It has been shown, that even a sparse AP set consisting only of df\_P0, F1 bandwidth, dA\_P0 and dA\_P1 achieves 84 % accuracy, when used on adult female /a/ sounds from the data corpus NASAL [Mal09].

In another study we investigated the necessity of P0 and P1 for nasality perception. We used an ordinary LPC of order 13 at 11025 kHz sampling rate on nasal and non-nasal speech. This low-order approach is just about able to capture the general formant structure, but not P0 or P1. Listening tests delivered a significant perceptual distance between nasal and non-nasal vowels. Thus, P0 and P1 are not necessarily the most prominent cue to nasality.

## Nasality in voice vs. musical sounds

We applied the knowledge to musical sounds in different studies. Near-field recordings from a Stradivari violin were post-processed to implement individual APs from Baken [Ker08]. Being asked on perceived changes to sound, test persons gave all kinds of explanations but did not mention nasality at all. This confirms again that combinations of APs will trigger perception of nasality rather than individual APs. In another study we boosted the signal by 3 dB, 6 dB or 10 dB in bands from 600 Hz to 1000 Hz, 600 Hz to 1500 Hz, and 900 Hz to 1500 Hz, corresponding to Dünnwald's definition. Again, after listening to six different musical pieces, no one mentioned nasality while reporting perceived changes.

Perception of nasal ingredients in musical sounds will be triggered by many possible AP combinations, but not necessarily by those agreed upon in the different fields of research. A violin resonance profile for instance, offers enough pole-zero combinations over a wide range to trigger nasality, and most of the energy is outside the low frequency focus of speech research. Another problem is that applicability of knowledge to musical sounds becomes

difficult when the pitch is higher than that of voice. Understanding nasality in musical sounds will finally request a likewise effort as for understanding nasal speech.

## Conclusions

There are different knowledge bases on acoustical properties (APs) for nasality perception in speech processing, in clinical research, and in musical acoustics. The most reliable APs found for nasality in speech do not translate to musical instruments, especially with high-pitch and multi-resonance sounds. In an honest listening test, the often cited Dünnwald definition for nasality cannot be confirmed. Knowledge-based modelling with a sparse AP set from the speech community, however, resulted in 80 % classification accuracy. Perceptual tests on nasality need very careful design, since results will largely be driven by language background.

## Acknowledgements

We thank Prof. R. Männer, University of Mannheim, for providing the data corpus NASAL, and we thank the BMBF for funding the violin project, reference no. AiF 1767X07.

## References

- [Bak02] Baken, R.J., Orlikoff, R.F.: *Clinical Measurement of Voice and Speech*, Singular Publications, 2002.
- [Bed82] Beddor, P. S., Strange, W., Cross language study of perception of the oralnasal distinction. *J. Acoust. Soc. Am.* 71 (6), 1551–1561, 1982.
- [Ber07] M. A. Berger, Measurement of vowel nasalization by multi-dimensional acoustic analysis, MSc thesis, Univ. of Rochester, NY, 2007.
- [Bog86] Bognar, E., Fujisaki, H., Analysis, synthesis and perception of French nasal vowels. In: *Proceedings of ICASSP*, pp. 1601–1604, 1986.
- [Che95] Chen, M. Y., Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers. *J. Acoust. Soc. Am.* 98 (5), 1995.
- [Del68] Delattre, P., Monnot, M., The role of duration in the identification of French nasal vowels. *International Review of Applied Linguistics* 6, 267–288, 1968.
- [Dic62] Dickson, D. R., Acoustic study of nasality. *J. of Speech and Hearing Research* 5 (2), 103–111, 1962.
- [Fan60] Fant, G., *Acoustic Theory of Speech Production*. Mouton, The Hague, Netherlands, 1960.
- [Fuj71] Fujimura, O., Lindqvist, J., Sweep tone measurements of vocal-tract characteristics. *J. Acoust. Soc. Am.* 49, 541–558, 1971.
- [Gla85] Glass, J. R., Zue, V. W., Detection of nasalized vowels in American English. In: *Proceedings of ICASSP*, pp. 1569–1572, 1985.
- [Hat58] Hattori, S., Yamamoto, K., Fujimura, O., Nasalization of vowels in relation to nasals. *J. Acoust. Soc. Am.* 30 (4), 267–274, 1958.
- [Haw85] Hawkins, S., Stevens, K. N., Acoustic and perceptual correlates of the nonnasal-nasal distinction for vowels. *J. Acoust. Soc. Am.* 77 (4), 1985.
- [Haj04] N. Hajro, Automated nasal feature extraction, MSc th., MIT, 2004.
- [Hei03] G. Heike, H. Dünnwald: Neuere Klanguntersuchungen an Geigen und ihre Beziehung zum Gesang, in *Festschrift Jobst Peter Fricke zum 65. Geburtstag*, System. Musikwissensch. Publ. by W. Auhagen et. al., 2003.
- [Hou56] House, A. S., Stevens, K. N., Analog studies of the nasalization of vowels. *J. of Speech and Hearing Disorders* 21 (2), 218–232, 1956.
- [Ker08] Kersten, J., Sprechen versus Singen - eine Klanganalyse an Musikinstrumenten, diploma thesis, faculty DMI, HAW, Hamburg, 2008.
- [Kra91] Krakow, R. A., Beddor, P. S., Coarticulation and the perception of nasality. In: *Proceedings of the 12th Int. Congr. of Phonetic Sciences*, 1991.
- [Lin61] Lintz, L. B., Sherman, D., Phonetic elements and perception of nasality. *Journal of Speech and Hearing Research* 4, 381–396, 1961.
- [Mae82] Maeda, S., Acoustic cues for vowel nasalization: A simulation study. *J. Acoust. Soc. Am.* 72 (S1), S102, 1982.
- [Mae93] Maeda, S., *Phonetics and Phonology: Nasals, Nasalization and the Velum*. Academic Press, Ch. Acoustics of vowel nasalization and articulatory shifts in French Nasal Vowels, pp. 147–167, 1993.
- [Mal09] Malhotra, I., Extraktion der Nasalität in Klängen, diploma thesis, faculty DMI, HAW, Hamburg, 2009.
- [Pru07] Pruthi, T., Analysis, Vocal-Tract Modeling and Automatic Detection of Vowel Nasalization, PhD diss., Univ. of Maryland, 2007.
- [Ste87] Stevens, et.al., Perception of vowel nasalization in VC contexts: A cross-language study. *J. Acoust. Soc. Am.* 82 (S1), p 119, 1987.
- [Zec02] Zecević, A., Ein sprachgestütztes Trainingssystem zur Evaluierung der Nasalität, Dissertation, Universität Mannheim, 2002.