Nasality in musical sounds – it is not a frequency band

Robert Mores

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

Introduction

A recent meeting of reputated violin researchers and luthiers, held in Cambridge, led to this review. While it seems that "nasality" is one of the frequently used and commonly understood terms, perceptual studies on nasality in sounds often conclude without significant results, including those presented in Cambridge. Automated measurements of nasal content in recorded violin sounds seem to remain unachievable.

The violin research community still trusts the early definition of Dünnwald, who did a tremendous work in measuring more than 1000 violins and defining four characteristic energy bands for violins. One of these bands he assigned the nasal band, and in the latest publication on this 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 also covers some own studies on capturing nasality.

Nasality in Speech Processing

A brief look into history shows that after a period of fragmented research the community settled with some well accepted APs. House and Stevens, searching for the ingredients of nasality in speech in 1956, found some prominence at 1kHz, an additional dip in the range between 700-1800Hz and a reduced A1, the amplitude of the F1, the 1st formant [Hou56]. In 1958, Hattori identified a resonance at 250Hz and a zero at 500Hz [Hat58]. Two years later, Fant confirmed the reduction of A1 and additionally noticed an increase of bandwidth of F1, F1BW, and an extra formant at 2 kHz, seen in form of a split 3rd formant [Fan60]. Especially such early results may have caused the violin research in the 70s to assume, that nasality is just a matter of extra energy somewhere in the range of 1 to 2 kHz. Such assumption can easily be forwarded with the argument that sinus resonances are simply an add-on to the oral resonances. In 1962 Dickson confirmed the contribution of an increased bandwidth for F1 but also for F2 [Dic62]. He also noticed an increase or decrease of amplitude and frequency of F1, F2 and F3. Fujimura and Lindqvist report the frequency-shift of F1 and an extra zero-pole around F1 [Fuj71]. One decade later Maeda observed a flattening of the spectra in the range of 300 to 2500 Hz, which clearly seems to contradict some of the earlier observations [Mae82]. Hawkins and Stevens concluded in 1985, that it was the degree of prominence of an extra pole around F1 that would most likely feature nasality [Haw85]. Bognar and Fujisaki, in their study on the four French nasal vowels in 1986, found an upward frequency shift of F3 and a downward shift of F2, resulting in widening of the F2-F3 region with two extra pole-zero pairs between 220 - 2150 Hz [Bog86]. Dang et al. linked observations in the spectrum with source features in 1994, when they assigned the lowest pole-zero pair to the maxillary sinuses [Dan94]. In summary, early research results are fragmented and do not encourage to build a general model.

Acoustical Properties	std0-1k	% extra poles	delta 1st - xpole	F1-FP0 & F1-FP1	A1-P0 & A1-P1	MFCC	F1BW & F1 profile	nPeaks 40dB	A1-H1
Glass 85	x	x	x						
Maeda 93				x					
Chen M. 95					x				
Hasegawa 04						x			
Pruthi 07	x			x	x		x	x	x
Chen N. 07					x				
our study				x	x		x		

Legend:	
std0-1k	std around center of mass 0-1kHz
% xpoles	min/max values of % of time there are extra poles at low frequencies
delta 1st - xpole	min/max values of differences between first pole and extra pole
F1-FP0, F1-FP1	frequency differences between F1 and extra poles P0 and P1
A1-P0, A1-P1	amplitude differences between F1 and extra poles P0 and P1
MFCC	mel frequency correlation coeff.
F1BW	bandwidth of F1
F1 profile	signal energy after passing a 100Hz band filter in relation to passing a 1kHz band filter, both centered around F1
nPeaks40dB	number of peaks within 40dB of the maximum amplitude in a spectral frame
A1-H1	amplitude difference of F1 and 1st harmonic H1, two methods

Table 1: Acoustical properties for capturing nasality in speech sounds, as preferred in the speech processing community

Table 1 gives an overview of today´s most well accepted APs for nasality in speech. Different sets of these APs are usually taken as knowledge-based parameters 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 bandwidth or amplitude for these extra poles. The respective APs introduced by Maeda and Chen M. have been reused until now [Mae93][Che95]. Pruthi has resolved many issues in his dissertation and 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 [Pru04]. He also changed the paradigm of static sinus resonance frequencies and identified the interdependence of nasal APs and vowel quality [Pru07]. The work of Chen N. is listed because it exemplarily demonstrates that the established APs are reused and that further accuracy progress is now expected by other, context-sensitive, measurements, i.e. the statistical difference in A1-P1 measurements for vowels with adjacent nasal consonants (NVN, NVC, and CVN) vs. vowels with no adjacent nasal consonants (CVC).

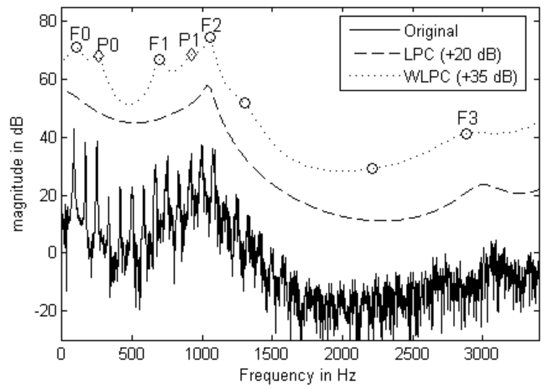

Figure 1: Nasal speech signal 'a' in the frequency domain and its related LPC and warped LPC spectrum

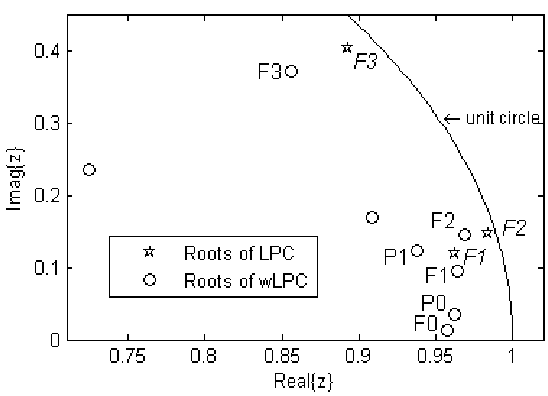

Figure 2: LPC and warped LPC coefficients of a nasal speech signal 'a' in the z-plane, for the related spectra see Fig. 1

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. However, reviewing the publications, it becomes clear, that the community is fully aware of the fact that there are many other APs around at higher frequencies and that earlier observations are true for individual test setups, however, with little chance for modelling. Pruthi has shown in his simulations that velum movement raises 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 [Pru05]. This confirms the complexity issue and explains the problem of generalisation for APs at the higher frequencies. In recent research, the speech community rather moves away from finding appropriate APs towards using additional cues. Examples are phonetic context and murmur thresholds or energy over time fluctuations which seem helpful to resolve the categorical question of nasality for further speech recognition improvements [Ber07] [Haj04] [Che07]. Such speaker-specific approach is reasonable for speech recognition when considering that inter-speaker variance within categories may be larger than intra-speaker categorical distances [Eng06].

Perceptual Issues

Apart from the analysis and modelling piece of work, the perceptual studies deliver a likewise heterogeneous scenario. House and Stevens in their study in1956 reduced A1 by 8 dB for the nasality response to reach the 50% level [Hou56]. Hattori et al. worked on poles and zeros in 1958 [Hat58]. Adding a pole around 250 Hz gave some perception of nasality, but adding the zero at 500 Hz did not, the combination of the two gave a much improved perception of nasality. Maeda confirmed in 1982 the importance of spectral flattening at low frequencies in producing the perception of nasality by listening tests [Mae82]. In 1985, Hawkins and Stevens inserted a pole-zero pair in the vicinity of the first formant, wider spacing of the pole-zero pair was found to be necessary for the perception of nasality [Haw85]. Bognar and Fujisaki studied the perception of French vowels in 1986, identifying the role of the formant shifts and of existing pole-zero pairs for phonemic and phonetic judgements [Bog86]. They also found biasing problems in using French native speakers to resolve perceptual questions on nasality. This brief review demonstrates that the method of varying only singular parameters within the complex multi-variant scenario fails to deliver the prominent APs.

Cross-language studies confirm the necessity of careful test design for perceptual studies. In their study in 1968, Delattre and Monnot presented stimuli to French and American English speaking listeners, differing only in vowel duration [Del68]. Shorter vowels were more likely identified as oral whereas longer vowels were more likely identified as nasal. Lintz and Sherman found in 1961, that the perceived nasality was less severe for syllables with a plosive environment than for syllables with a fricative environment [Lin61]. Beddor and Strange did not find consistent differences in responses from Hindi and American English speaking test groups when they investigated oral-nasal distinction in vowels in 1982 [Bed82]. However, these researchers identified that perception of oral-nasal vowel distinction is categorical for Hindi speakers, and more continuous for speakers of American English. Hawkins and Stevens did also not find significant differences between American English, Gujarati, Hindi and Bengali speaking test groups when they compared the 50% crossover points of the identification functions in 1985 [Haw85]. Stevens again identified similar responses to nasality content when working with Portugese, English and French speaking test groups in 1987 [Ste87]. However, British English speaking listeners preferred some murmur along with brief nasalization in the vowel, whereas French speaking listeners preferred a longer duration of nasalization in the vowel and gave little importance to the presence of murmur. Finally, Krakow and Beddor concluded in 1991 that nasal vowels presented in isolation or in oral context were more often correctly judged as nasal than when presented in the original nasal context, i.e. together with adjacent nasal consonants [Kra91]. In summary, even though there seem to be commonalities across languages, results from perceptual studies will still strongly depend on test group selection. For test persons with a language background containing phonemic nasalization, the trained categorical listening will be an obstacle to perceiving the degree of nasalization. And the results will also strongly depend on sound presentation, with or without phonemic context, oral or nasal context, duration of short-steady sounds, presentation with or without an onset, or, we could say with or without a plosive.

Clinical studies

Yet another perspective on nasality opens when reviewing clinical studies. Whereas the speech processing research aims at speech or speaker-specific feature extraction, clinical research aims at diagnosis and therapy of speech problems or inabilities. This different focus and context has led to other approaches in terms of analysis, modelling and data bases. Clearly, the Jones plane [Jon62], which represents vowel quality, or tongue position, will be shifted away from normal measures for children with Down´s syndrome, and nasality measurements are likely to fail for patients where the phonetic context is shifted due to conjoined cleft palates. In their clinical study in 2002, Baken and Orlikoff identified the following APs for nasality: larger F1BW, 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]. Some of these APs are similar to those found by the speech community, but in general these findings seem to stay behind state of the art. The classification study of Zecevic in 2002 aims at developing assisting tools for speech therapy [Zec02]. For classification with SVM, he decided to extract the first four formants in terms of frequency, amplitude and bandwidth on the basis of LPC (two levels of order), ignoring the extra poles P0 and P1. The investigated data corpus NASAL contains more than 3000 sounds from 116 male, female and infant speakers, following some guidelines of the Rinophoniebogen [Hep91]. Sound samples are classified into four nasal categories by speech therapists. This data corpus is particularly interesting for musical acoustics not only because of its differentiated classification in terms of nasal quality but also because of its emphasis on stand-alone vowels. Although some of the observations on individual changes to formants contradict those made by the speech community, the overall classification accuracy is well comparable with results in speech processing research.

Own studies between disciplines

In a brief study we used APs according to Table 1, but we extended the extraction method. F1, P0 and P1 were extracted using the warped LPC and a root solver on the LPC coefficients. Figure 1 clearly demonstrates the superiority of warped LPC against LPC when searching for properties on the low frequency side. Solving the roots of the LPC coefficients allows to identify measures for bandwidth and frequency of P0 and P1 even when these are masked by F1, see Figure 2. The model works without machine learning and achieves considerable classification accuracy. It has been shown that even a sparse AP set consisting only of F1-FP0, F1BW, A1-P0 and A1-P1 achieves 84 % accuracy, when used on adult female /a/ sounds from the data corpus NASAL, compared to the 94 % accuracy achieved in a 17 component AP set as suggested by Zečević, see Table 1 for abbreviations [Mal09].

In another study we investigated the necessity of P0 and P1 for perception. We used an ordinary LPC of order 13 (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. In listening tests the perceptual distance between nasal and non-nasal vowels was significant, even when presented with synthesized pitch. Therefore, P0 and P1 are not necessarily the prominent cue to nasality perception, even though the speech community agreed, that these APs are well extractable and sufficiently reliable.

Nasality in voice vs. musical sounds

In our approach of applying the knowledge to musical sound assessment we found both fields of research helpful. In an unpublished study in 2008, we post-processed near-field recordings from a Stradivari violin, implementing individual APs from Baken and Orlikoff [Ker08]. Being asked on any perceptible change of sound, test persons gave all kinds of explanations but did not mention nasality at all. This confirms again that combinations of APs rather than individual APs will trigger perception of nasality. In the same 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. This approach corresponds to the Dünnwald definition. Again, after listening to six different musical pieces, none of the test persons mentioned nasality while describing perceived changes.

Perception of nasal ingredients in musical sounds will be triggered by many possible AP combinations, but not necessarily 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. We have to admit that understanding nasality in musical sounds will finally request a likewise effort as for understanding nasal speech.

Conclusions

The knowledge base on acoustical properties (APs) for nasality perception seems to be stronger in the fields of speech processing or clinical research than in musical acoustics. This knowledge is not necessarily applicable to musical sounds. 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, phonetic context and duration of sound presentations.

Acknowledgements

We thank the research group around Prof. R. Männer, University 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, University of Rochester, Rochester, New York, 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), 2443–2453, 1995.

[Che07] N.F.Chen, Vowel nasalization in American english: acoustic variability due to phonetic context, ICPhS XVI Saarbrücken Germany 2007, ID1171, 905-908.

[Dan94] Dang, J., Honda, K., Suzuki, H., Morphological and acoustical analysis of the nasal and the paranasal cavities. J. Acoust. Soc. Am. 96 (4), 2088–2100, 1994.

[Del86] 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.

[Eng06] Engwall, O., Delvaux, V., & Metens, T. Interspeaker Variation in the Articulation of French Nasal Vowels. In *In Proceedings of the Seventh International Seminar on Speech Production* (pp. 3-10). Ubatuba, Sao Paolo, Brazil, 2006.

[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.

[Has04] Hasegawa-Johnson, et. al., Landmark-based speech recognition: Report of the 2004 Johns Hopkins summer workshop. Tech. rep., 2004.

[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), 1560–1575, 1985.

[Haj04] N. Hajro, Automated nasal feature extraction, MSc thesis, MIT, 2004.

[Hei03] G. Heike, H. Dünnwald: Neuere Klanguntersuchungen an Geigen und ihre Beziehung zum Gesang, in Festschrift Jobst Peter Fricke zum 65. Geburtstag, Systemische Musikwissenschaft published by Wolfgang Auhagen et. al., 2003.

[Hep91] Heppt, W.,et. al., Nasalanz – Ein neuer Begriff der objektiven Nasalitätsanalyse; Laryngo-Rhino-Otologie 70, pp. 169-228, 1991.

[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.

[Jon62] Jones, D.: An Outline of English Phonetics, W. Heffer & Sons Ltd. Cambridge, 9th edition, 1962.

[Ker08] Kersten, J., Sprechen versus Singen - eine Klanganalyse an Musikinstrumenten, diplome thesis, faculty DMI, HAW, Hamburg, 2008.

[Kra91] Krakow, R. A., Beddor, P. S., Coarticulation and the perception of nasality. In: Proceedings of the 12th International Congress of Phonetic Sciences. pp. 38–41, 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] I. Malhotra, Extraktion der Nasalität in Klängen – eine Werkzeugentwicklung unter MATLAB, diplome thesis, faculty DMI, HAW, Hamburg, 2009.

[Pru04] Pruthi, T., Espy-Wilson, C., Acoustic parameters for automatic detection of nasal manner. Speech Communication 43 (3), 225–239, 2004.

[Pru05] Pruthi, T., Espy-Wilson, C., Simulating and understanding the effects of velar coupling area on nasalized vowel spectra. J. Acoust. Soc. Am. 118 (3), 2024, 2005.

[Pru07] Pruthi, T., Analysis, Vocal-Tract Modeling and Automatic Detection of Vowel Nasalization, PhD dissertation, University of Maryland, 2007.

[Ste87] Stevens, K. N., Andrade, A., Viana, M. C., Perception of vowel nasalization in VC contexts: A cross-language study. J. Acoust. Soc. Am. 82 (S1), p 119, 1987.

[Zec02] Zečević, A., Ein sprachgestütztes Trainingssystem zur Evaluierung der Nasalität, Dissertation, Universität Mannheim, 2002.