On the use of accelerometer sensors to study nasalization in speech and singing voice

T Fux, A Amelot, L Crevier-Buchman, C Pillot-Loiseau, Martine Adda-Decker

On the use of accelerometer sensors to study nasalization in speech and singing voice

T. Fux[1], A. Amelot[1], L. Crevier-Buchman[1,2], C. Pillot-Loiseau[1], M. Adda-Decker[1]

[1]*Laboratoire de Phonétique et de Phonologie, CNRS, UMR7018, Paris, France*
[2]*Hôpital Européen Georges Pompidou, Univ. Paris-Descartes, Paris, France*
thibaut.fux@hotmail.fr

Abstract

This paper presents first results of a study aiming to explore data coming from nose mounted accelerometer during speech and singing tasks. One objective was to study the variations in the piezoelectric signal under variable speech and singing voice productions. Thus, only high-pitch and high-level singing are considered in this study. Four speakers (2 males, 2 females) produced isolated vowels, CVC and VCV non-words in nasal and non-nasal consonantal contexts. Our results suggest that the discrimination of nasal consonants remains possible in singing voice. A second part of this study investigates the correlation between acoustic and piezoelectric signals in vocalic sounds. A relative stable transfer function, with a frequency dip at low frequency around 500 Hz could be measured in our data. Results highlight a relative stable transfer function between audio and accelerometer signal for the vowels.

Keywords: accelerometer, nasalization, singing voice

1. Introduction

The proposed work is a pilot study which aim at investigating the usefulness of a nose mounted accelerometer to study its response in different contrastive conditions: oral *vs* nasal phonemes and speech *vs* singing. Nasal accelerometers have proven useful to locate a nasal sound production in a continuous spoken speech signal (Stevens, Kalikow, and Willemain 1975; Horii 1980; Lippman 1981; Brkan, Amelot, and Pillot-Loiseau 2012). The captured signal corresponds to skin vibrations during phonation which are related to the airflow through the nasal cavities and possibly to bone conduction. For example, the signal of the nose accelerometer enables the location of a nasal consonant in an oral vowel VCV context in modal speaking voice due to an increased intensity in the nasal consonant. Stevens, Kalikow, and Willemain (1975) found intensity differences of 10-20 dB between nasal consonants intensity and oral vowels intensity in the accelerometer signal but only 10 dB with high vowels such /i/. Even if the accelerometer method has already been explored for phonetic studies in speech, it remains important to study its behavior in singing voice. Do the observations done for speech still hold for singing voice? In particular, what happens if sound is produced with high intensity and relatively high pitch?

The spectral composition of the signal of a nose mounted accelerometer was studied by (Tronnier 1994; Tronnier 1998). Results show that the formant structure of oral vowels is not necessarily preserved in accelerometer data whereas the structure of nasal vowels remains almost unchanged. Trying to better understand the accelerometer signal intensity variations across different vowel types is a second part of this study aiming to explore the variation between audio spectrum and accelerometer spectrum. To this end, we propose to compute a transfer function as the ratio between the acoustic and accelerometer spectra in spoken and singing voice.

2. Database

To check the relevance of nose mounted accelerometers (also named contact microphone), which are piezoelectric sensors, for future studies of nasality and nasal vibrations in singers, we designed a set of production experiments making use of contrastive phonetic contexts. Subjects produced the designed material in both speaking (modal voice) and singing conditions. The collected data include speaking and singing of 4 non-professional singers (2 males, 2 females) recorded as follows. The subjects were placed in a soundproof room and asked to pronounce nine French oral vowels (/a/, /e/, /i/, /o/, /u/, /y/, /ɛ/, /œ/, /ø/) and 9 CVC and 9 VCV non-words composed of the three cardinal French vowels (/a/, /i/ and /u/), 2 nasal consonants (/m/ and /n/) and one occlusive consonant (/b/). Note that in a given word the two embedding consonants/vowels are identical. Recordings were carried out using a head mounted microphone (model C520L from AKG) and a nose mounted dual accelerometer i.e. model Twin spot from K&K sound. The accelerometer was connected to a pre-amplifier and recorded, simultaneously with the microphone at 44100 Hz. The accelerometer sensors were double taped on the nose at position 6 described in (Tronnier 1994). To further secure their positions surgical tape was used to fix wires on the cheeks. Both signals were then resampled at 16 kHz.

2.1. Database overview

Table 1 describes the captured data in terms of average F0 values, for the spoken and the sung material for the four subjects (F1, F2, M1, M2). The values are determined by considering only the isolated vowel productions in order to give an idea of the difference between our spoken and sung voices. Note that the production of non-words showed similar evolutions.

3. Measurements

For the presented experiments two kinds of measurement were made. The first one is the measurement of intensities of accelerometer signal for consonants and vowels. The second one is the microphone to accelerometer transfer function, i.e. the relation in the frequency domain between the audio spectrum and the piezoelectric spectrum.

Table 1: Average values of F0 for all subjects as measured on all isolated vowels. Bottom lines show F0 (in semitones) and Intensity differences between singing and spoken productions.

	F1	F2	M1	M2
$F0_{spoken}$ (Hz)	202	207	96	123
$F0_{singing}$ (Hz)	355	369	196	343
$\Delta F0$ (semitones)	9.8	10.0	12.3	17.7
ΔI (dB)	7.8	15.6	8.4	5.6

3.1. Intensity

The recorded material was manually segmented into C and V segments both for speech and singing voice. The intensity was measured using a 25 ms window length. Intensity measurements were carried out as follows: for the first (second, third) phoneme of the CVC/VCV patterns the measurements were carried out at 2/3 (1/2, 1/3 respectively) of their durations. Intensities of isolated vowels were measured in the middle of each segment (Larson and Hamlet 1987). We defined an intensity ratio (in linear scale) as the ratio between the intensities of a consonant and the neighboring oral vowel measured using the accelerometer signal. The presented intensity ratios are measured only for the first couple of consonant-vowel for each nonsense word (C *vs* V1 for V1CV2, and C1 *vs* V for C1VC2).

In fully oral VCV/CVC productions, the intensity of the accelerometer signal remains low both for vowel and consonant segments. If the VCV/CVC pattern includes a nasal consonant, the intensity tends to raise in the accelerometer signal, precisely in the region including the nasal consonant. An increase of the $\frac{I_C}{I_V}$ intensity ratio thus signals the presence of nasality or at least nasalization. In singing voice, vocal techniques may induce nasalization without the presence of nasal segments.

3.2. Transfer function

The transfer functions were measured using exclusively isolated vowels. The spectrum of the audio and the piezo signals were computed in the middle of the vowels using a *Blackman-Hanning* shaped window of 1024 points (64 ms). The principle of the transfer function measurement is to divide the output (piezo) by the input (audio) in the frequency domain with linear amplitude. However, periodic speech signals include low energy components between the harmonic peaks of the spectrum. This may generate a bias in the transfer function computation. A solution consists in splitting the spectrum into small frequency bands of harmonic and non harmonic regions. However, the size of the frequency band may be chosen in accordance to the F0 value. In this paper we preferred to use only the harmonic amplitude differences. Thus, the first step, for each vowel, is the estimation of F0 and then, using a peak picking algorithm, the location of harmonic peaks in the spectrum. We chose to limit this peak picking to the region between F0 and 3 kHz. Indeed, above 3 kHz the harmonics are often hard to identify due to low harmonic-to-noise ratios in the higher frequency regions of speech signals.

This harmonic peak picking has been done in both the audio and the piezo spectrum in order to compute the amplitude difference of their respective harmonics (i.e. the audio-to-piezo transfer function). This generated discrete transfer functions of various lengths related to the initial F0 values. Thus, to average multiple occurrences, transfer function curves were interpolated in the frequency domain to obtain the same number of points. This allows us to determine a mean transfer function for each subject and voiced production types (speech and singing) using all the isolated vowel occurrences.

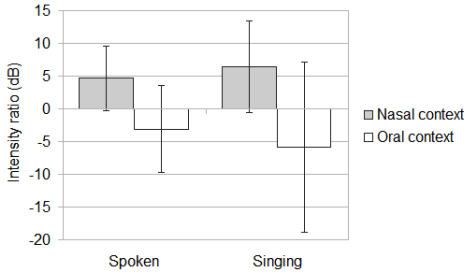

Figure 1: Average intensity ratio ($\frac{I_C}{I_V}$) between consonants (oral and nasal), and oral vowels for spoken and singing voice.

4. Results

This section presents various results evaluating the discrimination potential between nasal and oral productions, using an accelerometer sensor, in both the singing voice and speech. Furthermore, we also examined the frequency composition of the piezoelectric signal as compared to the audio.

4.1. Piezo Intensities: non-words, speech *vs* singing

Figure 1 shows average $\frac{I_C}{I_V}$ intensity ratios over the whole database. The figure compares two conditions: speech (left) and singing voice (right). Each condition gives contrastive results for nasal *vs* oral consonants. As expected, average results show positive intensity ratios for nasal consonants and negative ones for oral consonants in both speech and singing conditions with even stronger tendencies for singing voice. Results thus confirm the potential of accelerometers to locate nasality in singing voice. However, the high standard deviations in Figure 1 signal important variabilities. Future studies should better control for different factors giving rise to variation.

Figure 2 shows the average ratio values of the three repetitions merged according to the consonantal context and vowel type. The first observation is for the vowel contexts of /i/ and /u/ where amplitudes of nasal consonants are less salient. This is a well known and frequently observed effect, due to a larger increase of the accelerometer signal for high vowels than for low vowels (/a/). It may be related to a higher oral impedance and a larger surface of palate exposed to the airflow. This may favor the propagation of vibrations (Bundy and Zajac 2006; Gildersleeve-Neumann and Dalston 2001). Observations per speaker show that the differences between nasal consonants and vowels are not necessarily consistent among speakers between spoken and sung tasks. For subject F2, a semi professional singer, it is interesting to observe that higher ratios are obtained for the /a/ context in speech, but this result is reversed in singing voice. Similar variable results were found by Chen, Ma, and Yiu (2014) who measured more intense accelerometer signals for /i/ and /u/ than for /a/ for singers trained to resonant voice techniques.

Cases where the nasal consonants are less prominent than the neighboring vowels in piezoelectric signals could be observed. This may be explained by the fact that vowel acous-

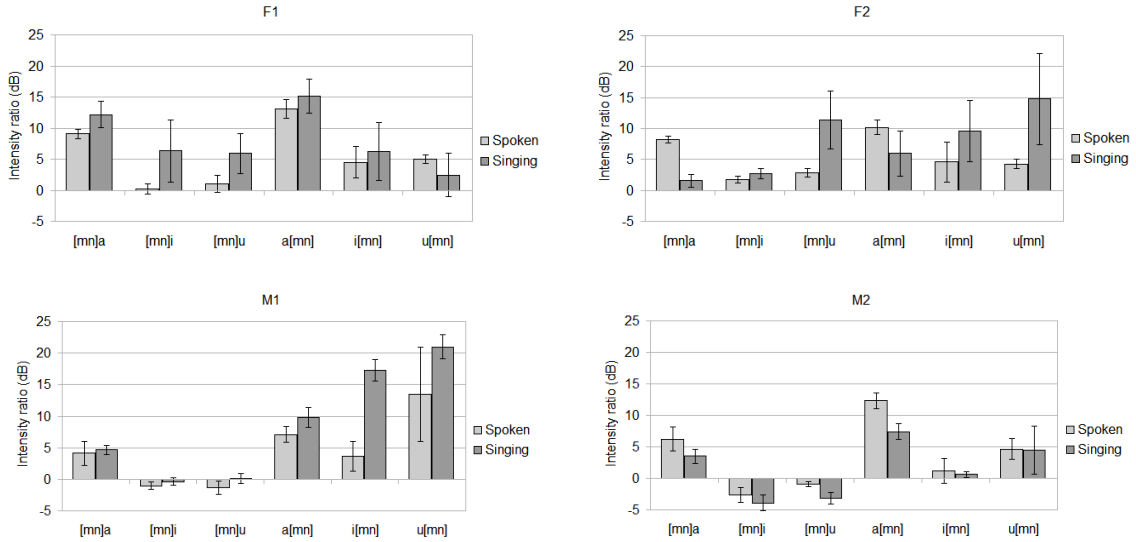

Figure 2: Mean accelerometer intensity ratios ($\frac{I_C}{I_V}$) between nasal consonants and oral vowels for the 3 repetitions of CVC and VCV patterns for each subject. [mn]V stands for CVC with m and n merged, V[mn] for VCV.

tic intensities generally increase more than those of consonants when moving from spoken to singing voice. When focusing on the isolated vowel material we could confirm that an acoustic intensity increase entails an increase in the piezoelectric intensity. However, acoustic/piezo intensity ratios are not constant with globally increasing production intensities. Results suggest that the piezoelectric signal is not only influenced by the overall production intensity, but that other factors may come into play. Among these are various articulation and voice production strategies adopted by the speaker or singer. It may also be related to the measurement method. To investigate these different hypotheses, future experiments will include a larger set of singers and a more controlled production protocol.

4.2. Audio-to-piezo transfer functions: toward a piezoelectric signal prediction for oral production.

In this section, we address the question of how piezoelectric signal intensities relate to the acoustic ones in different frequency bands. By comparing the spectral compositions of the piezoelectric and audio signals as a function of vowel type, we derive audio-to-piezo transfer functions which remain comparable for all the vowels (at least until 1.5 kHz). More interestingly, these transfer functions remain almost the same for spoken and singing tasks. Figure 3 shows transfer functions averaged over all isolated vowel productions for two speakers (F2 and M2). Similar shapes are obtained with a dip around 500 Hz and they remind transfer functions of bone conduction as shown in other studies (Won and Berger 2005). Obviously, transfer functions are speaker dependent and also depend on experimental parameters such as positioning and fixing of the sensors (accelerometer and microphone). Thus, piezoelectric intensity not only depends on the acoustic intensity but also on the spectral composition of the produced sound.

To further investigate this question, figure 4 shows three glissandi of /a/, /i/ and /u/ vowels produced by a semi-professional singer. The top panel corresponds to audio data, the bottom panel to piezo recordings. For each sensor, F0 and intensity (I) curves are provided. It is interesting to compare the relative shapes of F0 and I curves in piezo and audio data.

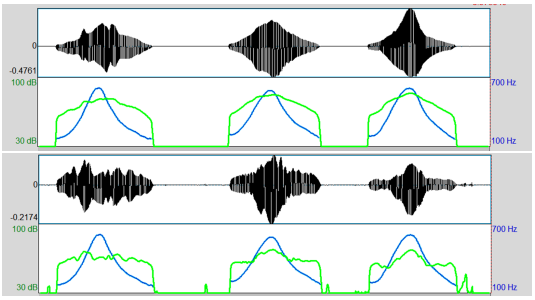

Figure 4: Glissandi signals for /a/ (left), /i/ (middle) and /u/ (right) with F0 (blue) and intensity (green) curves by subject F2 – audio (top) and piezo signal (bottom).

For /i/ and /u/ vowels, the I level at the beginning and the end of glissandi are relatively constant (not increasing) while audio intensity increases. In the middle of the /i/ and /u/ glissandi, the piezo I shape tends to follow the audio I shape. Different hypotheses may be proposed to explain this observation. A first hypothesis relates to the influence of harmonics and first formant positions around the "valley" zone of the transfer function, which in turn, influences the accelerometer intensity signal. Another hypothesis is related to the *passage* from production mechanism 1 to mechanism 2. To further explore these hypotheses, electroglottograph data and larger databases are needed.

5. Discussion

Accelerometer sensors are known to be useful for nasal production studies. If the objective is the localization of a nasal (or nasalized) sound in a continuous accelerometer signal, the main difficulty is the determination of a threshold to discriminate oral from nasal productions. Moreover, it appears to be difficult to predict the intensity of the piezoelectric signal for purely oral productions. High vowels are often considered to have higher piezoelectric intensity but our first results show no significant relation between vowel type and piezoelectric intensity.

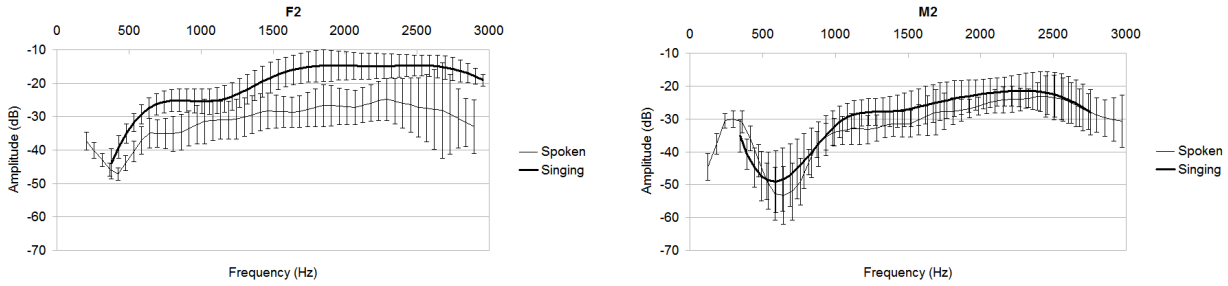

Figure 3: Transfer function computed as ratio between accelerometer signal and acoustic signal. The curves represent the mean transfer functions across all vowels produced by 2 speakers in spoken and singing voice (bold line).

Many points can be improved with respect to the recording procedure and must be tested in order to check their influence on the signal. First, the manner to position and fix the accelerometer may influence the frequency response of the sensor. A further bias may be due to velum and head movements, especially in singing. The use of multiple accelerometers to apply movement corrections can be envisioned. Finally, the accelerometer used in this study is a double accelerometer (one of each nose side) wired in parallel to obtain a single signal. If the positions of both the sensors are not exactly the same, the two accelerometers will not necessarily produce similar signals and will not necessarily be in phase. Furthermore, signal asymmetries may arise from asymmetric nose morphologies. Separate analyses of the two sensors should be studied to be sure there is no phase influence. A better control of all these factors should increase the reliability of the measurements and thus result in more reliable relations between accelerometer responses and oral productions. A more precise knowledge of this relation or transfer function will hopefully contribute to establish meaningful and robust thresholds for nasalization measurements and related studies.

6. Conclusion

The presented study investigated the use of accelerometer sensors in speech and singing, focusing on isolated vowels and nasal *vs* oral consonants in oral vowel contexts (CVC and VCV patterns). We made use of two types of measurements. First, an intensity $\frac{I_C}{I_V}$ ratio using the accelerometer signal was used to compare spoken and sung productions of the CVC and VCV patterns. Secondly, piezo *vs* audio spectrum intensity ratios were computed to produce corresponding transfer functions. Such transfer functions will contribute to give more precise interpretations of the observed variations in the piezo signals. Results concerning the $\frac{I_C}{I_V}$ ratio suggest that nose mounted accelerometers can be used in singing voice to locate nasal consonants since they remain more energetic than oral vowels (as nasal consonants also tend to do in speech). Secondly, this results of the piezo/acoustic spectrum ratios (or transfer functions) suggest that the spectral composition, in particular the phonetic type of the sound, affects its piezo intensity and that this should be taken in account when using a piezoelectric intensity normalization method such as the HONC method (Horii 1980). Some authors suggested a piezoelectric intensity F1 dependence, explained by physiological phenomena (Stevens, Nickerson, et al. 1976). The achieved results are consistent with this finding, and further suggest an F1 dependence with the sensor frequency response.

7. Acknowledgments

This work was partially funded by the European FP7 i-Treasures project (Intangible Treasures - Capturing the Intangible Cultural Heritage and Learning the Rare Know-How of Living Human Treasures FP7-ICT-2011-9-600676-i-Treasures). It was also supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

8. References

Brkan, A., A. Amelot, and C. Pillot-Loiseau (2012). "Utilisation d'un accéléromètre piezoélectrique pour l'étude de la nasalité du Français Langue Etrangère". In: *Procceedings of JEP-TALN-RECITAL*, 689–696.

Bundy, E. L. and D. J. Zajac (2006). "Estimation of transpalatal nasalance during production of voiced stop consonants by noncleft speakers using an oral-nasal mask". In: *Cleft Palate – Craniofacial Journal* 43, pp. 691–701.

Chen, Fei C., Estella P.-M. Ma, and Edwin M.-L. Yiu (2014). "Facial Bone Vibration In Resonant Voice Production". In: *J. of voice*.

Gildersleeve-Neumann, C. E. and R. M. Dalston (2001). "Nasalance scores in noncleft individuals: Why not zero?" In: *Cleft Palate – Craniofacial Journal* 38, pp. 106–111.

Horii, Y. (1980). "An accelerometric Approach to Nasality Measurement: a preliminary report". In: *Cleft Palate Journal* 17, pp. 254–261.

Larson, P. L. and S. L. Hamlet (1987). "Coarticulation effects on the nasalization of vowels using Nasal/Voice amplitude ration instrumentation". In: *Cleft Palate Journal* 24, pp. 286–290.

Lippman, R. P. (1981). "Detecting nasalization using a low cost miniature accelerometer". In: *J. of Speech and Hearing Research* 24, pp. 314–317.

Stevens, K. N., D. N. Kalikow, and T. R. Willemain (1975). "A miniature accelerometer for detecting glottal waveforms and nasalization". In: *J. of Speech and Hearing Research* 18, pp. 594–599.

Stevens, K. N., R. S. Nickerson, A. Boothroyd, and A. M. Rollins (1976). "Assessment of Nasalization in the Speech of Deaf Children". In: *J. of Speech and Hearing Research* 19, pp. 393–416.

Tronnier, M. (1994). "Tracing Nasality with the Help of the Spectrum of a Nasal Signal". In: *Australian Conference on Speech Science and Technology*, pp. 330–335.

— (1998). "Nasals and Nasalisation in Speech Production with Special Emphasis on Methodology and Osaka Japanese". eng. PhD thesis. Lund University, p. 220.

Won, S. Y. and J. Berger (2005). "Estimating transfer function from air to bone conduction using singing voice". In: *Proceedings of the International Computer Music Conference*, pp. 123–126.