Digestive activity evaluation by multi-channel abdominal sounds analysis

R. Ranta, *Member, IEEE,* V. Louis Dorr, Ch. Heinrich, *Member, IEEE,* D. Wolf, F. Guillemin

Abstract

This paper introduces a complete methodology for abdominal sounds analysis, from signal acquisition to statistical data analysis. The goal is to evaluate if and how phonoenterograms can be used to detect different functioning modes of the normal gastrointestinal tract, both in terms of localization and of time evolution during the digestion. After the description of the acquisition protocol and the employed instrumentation, several signal processing steps are presented: wavelet denoising and segmentation, artifact suppression and source localization. Next, several physiological features are extracted from the processed signals issued from a data-base of 14 healthy volunteers, recorded during 3 hours after a standardized meal. Data analysis is performed using a multi-factorial statistical method. Based on the introduced approach, we show that the abdominal regions of healthy volunteers present statistically significant phonoenterographic characteristics, which evolve differently during the normal digestion. The most significant feature allowing to distinguish regions and time differences is the number of recorded sounds, but important information is also carried by sound amplitudes, frequencies and durations. Depending on the considered feature, the sounds produced by different abdominal regions (especially stomach, ileo-caecal and lower abdomen regions) present a specific distribution over space and time. This information, statistically validated, is usable in further studies as a comparison term with other normal or pathological conditions.

I. Introduction

One of the oldest means of physiological investigation, still currently used in clinical routine, is the auscultation. The instrumentation is simple (stethoscope) and its utility is largely recognized especially for

R. Ranta, V. Louis-Dorr and D. Wolf are with the Centre de Recherche en Automatique de Nancy (CRAN), Nancy-University, CNRS, 2 avenue de la Forêt de Haye F-54516 Vandoeuvre-les-Nancy; e-mail: radu.ranta@ensem.inpl-nancy.fr

C. Heinrich is with the Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection – LSIIT UMR 7005, University of Strasbourg, BP10413 67412 Illkirch-Graffenstaden, France.

F. Guillemin is with the Centre de Recherche en Automatique de Nancy (CRAN), Centre de Lutte contre le Cancer (CAV), Nancy-University, CNRS, av. de Bourgogne, F-54516 Vandoeuvre-les-Nancy

cardiac and pulmonary sounds. Relatively little studied for abdominal sounds, although the first papers appeared a century ago [1], the phonoenterography presents significant potentialities [1–8]. Different applications can be imagined, from the study of the normal physiology (classifying digestion phases or abdominal regions) to clinical routine (functional diseases diagnostic aid, post-surgical monitoring) or pharmacological research (medication effect on the gastro-intestinal activity).

Two types of approaches for phonoenterogram interpretation are proposed in the literature: the first one focuses on the analysis of individual sounds, classified for example in first, second and third degree sounds [9], clicks, multiclicks and complexes [3], intestinal bursts and regularly sustained sounds [10]. The second approach, adopted here, is more widely used and tries to analyze sequences of phonoenterograms using different, so called, "activity indices" like mean sound duration, silence between sounds duration, signal energy, and so on. The underlying hypothesis is that abdominal sounds, *recorded upon long periods of time* and *in several abdominal locations*, are representative of the physiological activity, either normal or pathological. Under this hypothesis, changes in the patient status are reflected in the activity indices, which can be used to assess and quantify differences among normal or pathological physiological states [2, 5, 6, 11–13]. However, phonoenterogram interpretation is particularly difficult: there is no consensus on a method for processing and analyzing abdominal sounds over long durations and in simultaneous locations (see the comparative studies [14] or [15]).

The first goal of this paper is to propose a complete (although not unique) signal acquisition, processing and analysis methodology, able to extract significant activity indices from long-term multi-channel phonoenterograms. The second and more important goal is to use these indices to analyze the influence of the physiological factors, such as the patient, the abdominal region and the digestion phase, on the phonoenterograms. A detailed statistical analysis is performed to check if phonoenterograms, characterized by simple physiological activity indices, can be used to make statistically valid affirmations about the digestion: which regions can be distinguished, when, how much time one has to record (listen) and which are the indices that can be used to make this difference.

We focus here only on the digestion healthy volunteers, in order to obtain a phonoenterographic point of view description of the normal functioning of the digestive tract, in the given controlled recording conditions. The obtained spatial-temporal (regions–digestion phases) distribution of the abdominal activity is statistically validated through a non-parametric factorial data analysis (ANOVA type) and can constitute a comparison term for other normal or pathological phonoenterogram data, recorded and extracted in similar conditions.

After this introduction, the second section describes the signal acquisition and processing methods.

The main addressed issues are the protocol used to ensure the repeatability of the measures and the novelties introduced in the denoising procedure. The next section presents the experimental design and the statistical analysis methods. The fourth section details and discusses the results and is followed by a conclusion and by possible future research directions.

II. Signal Acquisition and Processing

The current paper uses similar acquisition protocol and instrumentation as our previous works [12, 13, 16–18], so we focus on the test protocol used to ensure the repeatability of the measures. Moreover, as most of the employed signal processing methods were already introduced and validated in the cited papers, only a brief reminder is presented here.

A. Rationale

Regardless of the data analysis methodology, the obtained results are influenced by the elements of the acquisition and signal processing chain. To reduce the possible variations of the activity indices due to the acquisition[1], we place ourselves in a controlled environment (identical instrumentation, recording conditions and signal processing for all recording channels and all patients). These conditions must of course deal with phonoenterogram difficulties:

- Long-term recordings have a high variability, either in time (digestion phases), in location (abdominal region) or among patients. For a valid interpretation, it is necessary to record on several patients, in different abdominal locations and using similar recording conditions (standardized protocol and instrumentation).
- Individual abdominal sounds have a highly irregular character and random appearance (although quasi-periodic bursts have already been detected by [1] a century ago), and they are contaminated by noise and artifacts (movements, heart beats, ...). It is then necessary to detect, segment and denoise them, as well as to characterize them in order to eliminate the artifacts.

B. Data Acquisition

Experimental Protocol. Our data-base consists of 14 healthy volunteers of medium build, partially based upon the data-set used in [13][2]. All phonoenterograms were recorded in similar conditions and, to avoid

[1]Real clinical auscultation conditions can be highly variable and they could influence the results of the analysis.

[2]Compared to [13], some patients were eliminated and others were added to obtain an homogeneous data-base.

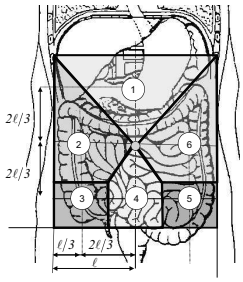


Fig. 1. Stethoscope placement and abdominal regions: ℓ stands for the distance between the navel and the lateral side

perturbing the normal digestion of the volunteers, the choice of the meal was adapted to their alimentary habits: a standardized breakfast was taken at about 8:30 a.m. and consisted in a cup of tea/coffee, 2 bread rolls, 200 ml of orange juice and 1 yoghurt. The end of the recording, almost 3 hours later (168 minutes), is very close to the main meal of the day, taken at about 12:00, so we consider that we can follow the digestion (post-prandial period) and have the "hungry" period (pre-prandial) for each volunteer. A completely lied-down position (which could be more appropriate for good recording conditions) was difficult to accept and maintain for the whole recording duration, so the volunteers' position was halfway between lying and sitting so they could watch television (using a headset to avoid phonoenterogram contamination by the TV sound). In order to minimize movement artifacts, they were instructed not to change their position during the recording.

Acquisition Sites. To obtain local information for different abdominal regions, six recording channels (1,2, ...6) were used (see Fig. 1). In order to respect inter-patient variability, channels 1, 2, 4 and 6 were positioned at equal distances from the navel. This distance was taken at 2/3 from the total distance between the navel and the lateral side of the abdomen. Channels 3 and 5 were aligned to channels 2, 4 and 6. These positions aim to record sounds over significant abdominal structures and to cover the whole abdominal surface (1 on duodenum; 2 on the ascending colon; 3 on the ileo-caecal valve; 4 on the small bowel; 5 and 6 on the descending colon).

Acquisition Instrumentation. Most of the literature reports classic microphones for acquiring abdominal sounds (as for example in [1, 2, 4, 5, 19]). Commercial electronic stethoscopes were used by Craine *et al.* [6, 7]. We have followed the approach of Garner and Ehrenreich [3], who adapted electret microphones to classic stethoscope heads. The six sensors were fixed on the abdomen wall with an elastic band.

Obviously, to these sensors we associated conventional analog electronics containing adjustable voltage amplifiers and band-pass anti-aliasing filters, calibrated to the bandwidth of the AD converter (Nicolet®

Vision 8 channels digital acquisition system, 16 bits resolution). According to the frequency characteristics of the signal (band-limited between 100 and 500 Hz, see next paragraph), the chosen sampling frequency was 5000 Hz.

The advantage of this instrumentation choice is that it allows a perfect control of the acquisition parameters: the six channels were calibrated in the same manner, and this calibration was verified before each recording.

The developed acquisition system has its pitfalls: first, the frequency response due to the stethoscope head is not flat. Still, commercial electronic stethoscopes also present selective frequency responses, which vary considerably from one to another [20]. Second and more important, the six channels might have different responses among regions and among volunteers, because of the variations in the fixation system. With this ideas in mind, we tried to evaluate, for one hand, the frequency response in the frequency band of interest and, on the other hand, the influence of the pressure applied by the stethoscope head against the patient's abdominal wall.

The frequency responses of the sensor were evaluated for different pressures, situated in a large enough interval ($\approx$400 Pa to $\approx$3500 Pa) to cover the actual pressures of the stethoscope head on the abdominal wall during the phonoenterogram recording. Measures were done in an anechoidal chamber by pressing the stethoscope head against an abdomen phantom using different force values and a calibrated white noise source between 100 and 1000 Hz. As expected, the frequency response of the stethoscope head is not flat, unlike the frequency response of the microphone alone (compared to those recorded using only the microphone, all sounds acquired through a stethoscope head are amplified from 5 to 30 dB over the band of interest [100-500] Hz, see Fig. 2). The curves presented in Fig. 2 correspond to different pressures and, as it can be seen, they are rather slightly influenced by the pressure in the frequency band 100-500 Hz, although more important variations appear for higher frequencies (around 800 Hz).

Finally, a last evaluation of the acquisition system was performed by medical expertise. Several minutes of the recorded phonoenterograms were listened and annotated by a clinician, who confirmed the good quality of the recordings from the medical interpretation point of view.

We conclude therefore that the frequency response of the instrumentation does not distort the physiological information carried by the abdominal sounds. Moreover, it does not significantly vary because of the fixation system, and thus neither among different recordings (in time or across the different regions or different patients). Therefore, no acquisition bias affects the signal analysis and presented in the sequel, and comparisons among regions and patients make sense, as the recordings were performed in similar conditions.

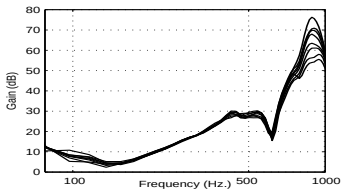


Fig. 2. Superimposed frequency responses for different sensor pressures varying from 390 to 3510 Pa (with a step of 390 Pa), corresponding to masses between 0.05 and 0.4 kg (0.05 kg step) placed on the stethoscope head

Signal. Healthy phonoenterograms are characterized by a succession of isolated short events. The signal consists of a sparse succession of non-stationary impulsive sounds (Fig. 3a). They can appear in periodic bursts of activity (3 to 12 per minute, according to the place and time of their generation) [1, 2]. The parts of the signal which separate the sounds, called in the bibliography "periods of silence", are not actually completely quiet. Noise due to acoustic effects of the stethoscope and to other low frequency physiological sounds (breathing, blood flow) is superimposed to the informative signal and must be taken into account in any further processing.

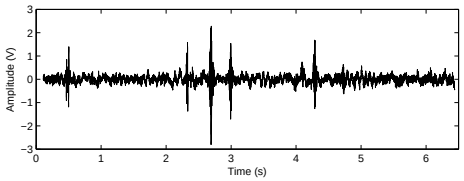
The frequency content of the phonoenterogram is relatively poor. The literature indicates maximum frequencies of the abdominal sounds lower than 1000-1500 Hz [4, 5, 7, 21], even if other values are mentioned (5000 Hz for example in [3]). The principal frequency of the abdominal sounds is generally higher than the frequencies of the cardiac and pulmonary sounds and sometimes a high-pass filtering at 80 Hz is used to eliminate the influence of the latter [5]. The frequency content of the noise is almost identical to that of the signal and it cannot be eliminated by simple filtering.

The literature description of the abdominal sounds is confirmed by our observations. Their frequency content is band-limited: only approximately 0.5% of the signal energy is located beyond 1000 Hz and only approximately 2% beyond 500 Hz. In fact, almost all of the phonoenterogram energy is situated between 100 and 500 Hz (Fig. 3b).

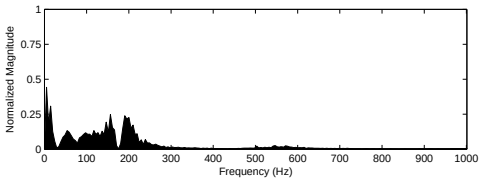
C. Denoising and Segmentation: $HystD$ Algorithm

Considering the signal characteristics (sparse transients), non-stationary denoising algorithms seem to be the most appropriate. This hypothesis was validated by our previous work and by several other authors, who developed wavelet-based algorithms for abdominal sounds denoising [13, 22–26]. Automatic segmentation procedures are proposed in [13, 19, 24].

The first wavelet-denoising algorithm applied on bowel-sounds was the iterative $WTST$-NST pro-



(a) time course



(b) spectral content (normalized to unitary energy)

Fig. 3. Typical phonoenterogram recording

posed by [23] and derived from the "peeling algorithm" of [27]. In [16, 18], we have shown that $WTST$-NST may be seen as a fixed-point algorithm and we have analyzed and determined its convergence conditions by introducing generalized Gaussian (GG) modelling of the wavelet coefficients. This approach leads to a "minimal denoising" algorithm $MinD$, completely parameter-free and ensuring a maximum information extraction from the measured signal. The main drawback of these methods is the fact that they tend either to over-estimate the number of individual abdominal sounds ($WTST$-NST and $MinD$) or to distort the detected ones (especially $WTST$-NST with a different parametrization $F_a > 3$, see [26]).

Several improvements were proposed in the last years. Fractal dimension estimation in the wavelet domain was used by [25, 28] to diminish the distortion of the detected events. Wiener filtering (in the wavelet domain also) was proposed by [26] to avoid over-detection and minimize distortion. In [13] we proposed a different approach, called "hysteresis denoising", which achieves denoising and segmentation in the same time, ensuring also a limited distortion of the segmented events. A slightly modified version of this algorithm (5) is briefly reminded here. Comparison with other recent signal processing developments, such as those introduced by [10, 19, 25, 26, 28], is beyond the purpose of this paper.

We consider the model $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{z}$ is the noisy discrete-time signal of length N, $\boldsymbol{x}$ is the noise-free unknown version of $\boldsymbol{z}$ and $\boldsymbol{n}$ the noise. Synthetically, the orthogonal discrete wavelet transform (DWT) of $\boldsymbol{z}$ writes:

$$\boldsymbol{z} = \sum_{p,j} w_z^{j,p}\ \boldsymbol{\psi}^{j,p} + \sum_{p} w_z^{M,p}\ \boldsymbol{\phi}^{M,p}, \tag{1}$$

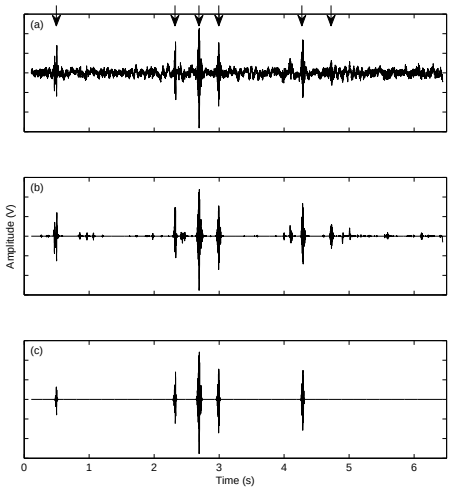


Fig. 4. (a) Example of 6.5 seconds of phonoenterogram, (b) its denoised version using $WTST$-NST ($F_a = 3$) and (c) novel $HystD$ thresholding. Expert identified events are indicated by arrows

where $j = [1 \dots M]$ is the scale, $p = [1 \dots 2^{-M}N]$ the position, ψ the wavelet, ϕ the scaling function and M the analysis depth [29]. The denoising threshold is computed using the fixed-point iteration:

$$T_{j,k+1} = F_a \sqrt{\frac{1}{N} \sum_p \left(w_z^{j,p} \max\left(0, \text{sign}(|w_z^{j,p}| - T_{j,k})\right)\right)^2}, \tag{2}$$

with $T_{j,k}$ the threshold at scale j, iteration k and F_a a multiplicative constant. This constant is user chosen for $WTST$-NST or function of the estimated Generalized Gaussian (GG) probability law of shape u followed by the wavelet coefficients for $MinD$ (see [18] for the proof):

$$F_{am} = \sqrt{\frac{3\Gamma(\frac{1}{u})}{u}(ue)^{\frac{1}{u}}}, \tag{3}$$

with $\Gamma(t) = \int_0^\infty e^{-x}x^{t-1}dx$. This F_{am} constant (subscript m stands for minimal) insures the convergence of the algorithm to a non-null threshold having a low value and thus leading to a maximal information extraction from the measured signal (minimal distortion).

For the implementation, the shape parameter u was estimated using the absolute empirical moments m_1 and m_2, (with $m_r = \mathbb{E}[|z|^r]$, see [30, 31]).

By slightly modifying the proposition made in [13] we introduce a correction term for F_{am}, leading to a new constant:

$$F_{ao} = \max(F_{am}, K_c F_{am}), \quad \text{with} \tag{4}$$

$$K_c = \sqrt{\frac{4 \log N}{3\sqrt{2\pi e}}}. \tag{5}$$

It is easily verified that for $u = 2$ (Gaussian law), $F_{ao} = \sqrt{2 \log N}$, which is the well known universal threshold proposed by [32].

The rationale behind this modification is the following: the universal threshold is constructed to asymptotically eliminate all the (Gaussian) noise from the measured signal. To adapt it to our iterative framework and to the GG case, we propose to modify the F_{am} constant, the goal being to achieve comparable performances of noise elimination. Therefore, an iterative algorithm (2) using the F_{ao} constant (4) will lead to a high threshold value and thus a quasi-complete noise cancelling. Moreover, applying this threshold on the approximation scale of the wavelet decomposition (low frequency), an approximate segmentation of the signal or, more precisely, a detection of the impulsive abdominal sounds, is also obtained. The price to pay for this high threshold is the distortion of the detected sounds (Figs. 4 and 5). It is then natural to combine the two iterative algorithms in one hysteresis denoising algorithm $HystD$: a high threshold (obtained with F_{ao}) to detect the greatest coefficients of each scale, and a low threshold (obtained with F_{am}) to select the "big enough" coefficients located in the neighborhood of those selected by F_{ao}.

A similar method was successfully applied in [13], so we only give some illustrative examples in Fig. 4 and 5. As it can be seen in Fig. 4, the $HystD$ algorithm detects almost all of the individual abdominal sounds (the sixth one, undetected, was hardly hearable by the expert, but it was confirmed after several listenings). Classic iterative $WTST$-NST detected it, but it also detected many parasite sounds which, on one hand, confused the expert and, on the other hand, made the segmentation almost impossible (and thus also the artifact elimination and the multi-channel processing, which are partly based on the characteristics of the segmented sounds). Nevertheless eliminating some of the sounds (as long as it remains marginal and similar for all recordings), does not influence the relative comparisons among regions and time sequences, although it might shift the absolute values.

Another situation can be observed for the first detected sound, which visually seems distorted. A close examination of its time course and spectral content reveals that the denoising/segmentation procedure eliminated the low frequency components (below 30 Hz in fact), which are quite energetic. The auditive impression confirms this analysis: almost no difference can be heard among the three denoised versions and the original sound, except for the background noise (original signal) or a succession of parasite clicks ($WTST$-NST).

Although not presented in Fig. 4, the iterative thresholding using F_{ao} (4), as well as the universal thresholding, furnished denoised estimated signals visually similar to the one obtained by $HystD$ (easy to segment and thus facilitating the number of sounds counting). Still, a detailed examination of the result

(see Fig. 5 for a zoom on the third detected abdominal sound in the example) shows that physiological characteristics of the sounds (amplitude, frequency, duration) might be distorted by a too high threshold. Therefore, having in mind that our goal is to extract activity indices from long-time recordings, a good compromise was offered by the $HystD$ algorithm.

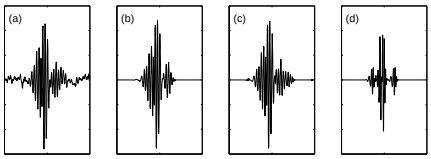


Fig. 5. Distortion comparison: (a) Example of an abdominal sound, (b) its denoised version using $WTST$-NST ($F_a = 3$), (c) using $HystD$ and (d) using an iterative fixed-point algorithm (2) with F_{ao} (4)

D. Artifact Elimination and Multi-channel Processing

Artifact elimination. Before proceeding to the multi-channel processing, we have decided to heuristically eliminate remaining artifacts. Although less sensitive than $WTST$-NST, $HystD$ (as certainly no denoising method) cannot ensure complete elimination of undesirable perturbations. In fact, noise cancelling methods deal with stationary noise, but non-informative (from a phonoenterographic point of view) signals are not treated. Indeed, heart beats, patient movements, cough, can be considered as informative events by the wavelet denoising/segmentation algorithm, and this kind of events are unfortunately unavoidable in long time measurements. We have introduced therefore *a priori* knowledge at this stage of our abdominal sound processing method. For each segmented event, we have computed the most popular physical features: the duration, the energy and the frequency spectrum, as in [4–7]. To avoid windowing effects, these characteristics were computed from the wavelet decomposition: duration as the union of the wavelet supports, the energy as the squared sum of the wavelet coefficients and the frequency spectrum as the sum of the wavelet spectra. The events which did not fit the literature description of an abdominal sound were eliminated: sounds shorter than 20 ms (like hair and skin friction on the stethoscope membrane), longer than 5 seconds (movements) or having more than half of their energy below 80 Hz (like heart beats and respiration). On real signals, several tests showed that this approach was more effective than high pass filtering or even adaptive filtering (with a reference stethoscope placed on the chest, for heart, cough or movement detection).

Multi-channel processing. There are two steps of multi-channel processing (see Fig. 1 for stethoscope placement). The first one concerns artifact elimination by cross-validation. In fact, we assumed that

real abdominal sounds propagate inside the abdomen. Therefore, we have eliminated all sounds that are not quasi-simultaneously acquired by at least two stethoscopes (that is, their time supports are strictly disjoint).

The second step is the localization technique. We have discussed different methods in [17]. In fact, very few publications present a multi-channel approach, and most of those who do it (like [5] for example) treat the recordings in a completely independent and parallel manner: they quantify abdominal sounds independently for each abdominal region and no propagation is taken into account. Two approaches were proposed by Craine *et al.* [7], who perform source localization by triangulation, and by Dimoulas *et al.* [19], who use a more elaborated decision tree adapted to their hierarchical segmentation technique. The first one is inaccurate, because of the high anisotropy of the propagation environment while the second one is inapplicable in our detection and segmentation framework. In fact, the technique we proposed in [17] reduces to what Dimoulas *et al.* call CPA (Closest Point of Approach [19]): indeed, the simplest hypothesis and, in the actual state of knowledge, the most accurate, is that the recorded sounds are louder when the stethoscope is placed closer to their origin. Therefore, we have proposed a 6-region partition of the abdomen, as indicated in Fig. 1. For each sound, we check its maximum amplitude (acoustic intensity) on each of the stethoscopes that acquired it, and its origin is placed inside the region indicated by the greatest value.

III. Data Analysis Method

Our first proposals for long-term monitoring, based on a set of activity indices, were made in [12, 13]. This paper proposes a different phonoenterogram characterization, directly based on the most significant physiological indices instead of principal components. Therefore, the findings are directly exploitable by clinicians: median values for several physiological indices are given for different abdominal regions and for different post-prandial time intervals.

A. Feature Extraction

Several empirical activity indices are proposed in the literature: number of sounds by time interval [1, 5, 7, 11, 33–35], the sounds duration, energy, power or amplitude integrated over a period of auscultation [2, 3, 5, 6, 33–35], sounds main frequency [5, 34], silence between sounds duration (integrated or averaged over time) [6, 7, 33, 35]. The most commonly considered time interval is the minute, which corresponds to the range of clinical auscultation duration by region. Summarizing, we have considered in our study nine activity indices, evaluated for each channel and for each minute of recording: the number of sounds

(N_m), the total energy (E_m), the total duration of sounds (D_m), the median energy of sounds (E_μ), their median duration (D_μ), their median power (P_μ), their median main frequency (f_μ), their median acoustic intensity (amplitude) (I_μ) and the median duration of silence periods between sounds ($D_{s,\mu}$).

Each minute of recording can then be represented as a point in the nine-dimensional space obtained from the nine activity indices. Considering our data-base, we have 14112 such points, representing 168 minutes for each of the 6 regions of the 14 patients. Interpreting all this information reveals to be difficult because of the high dimension of the representation space and, furthermore, because of the probable redundancy of the nine features[3]. To diminish the variable redundancy and thus the dimension of the representation space, we propose a guided feature selection step based on a correlation/ PCA analysis: after PCA, the first four principal components c_1 to c_4 were retained, as they describe more than 80% of the variance. Next, instead of projecting the data onto the reduced principal component space, the original features which are the most correlated with these first 4 principal components were retained: I_μ, N_m, f_μ and D_μ (correlated respectively to c_1 to c_4 at least 0.7). We discarded redundant variables as D_m, highly correlated with N_m (0.88), E_μ and P_μ, highly correlated between them (0.80) and with I_μ (0.67, respectively 0.68), E_m and $D_{s,\mu}$, uncorrelated with any of the principal components. The first three retained features are almost orthogonal (the maximum correlation coefficient among them is smaller than 0.2), while the last one (D_μ) is a little more correlated (0.4 with I_μ)[4]. This approach permits an easier comparison with the literature and, above all, a natural physiologic interpretation.

B. Statistical Data Analysis

The basic hypothesis is that all recordings are acquired in similar conditions, i.e., after a standardized meal and from a normal population, without particular digestion types (diseases, nutritional habits). The obtained analysis data-base consists of $14 \times 6 \times 168 = 14112$ points (minutes of phonoenterogram) in a four-dimensional feature space (I_μ, N_m, f_μ and D_μ).

We recall that the aim of this paper is to determine if the processed phonoenterograms can be used to evaluate differences among different physiologic conditions (digestion evolution) and/or recording sites (abdominal organs). A first visual analysis can be done by plotting the median values of the 4 retained variables (over the 14 patients), computed for every minute and every region (Fig. 6): the displayed

[3]In [12, 13] we have proposed principal component analysis (PCA) to reduce the number of retained features and to decorrelate them.

[4]All the $p-values$ for the correlation coefficients presented here are highly significant ≈ 0. This is consistent with the very high amount of data, as for every variable we have more than 14000 measures.

graphs seem to indicate differences among regions and time evolution during the digestion, but these differences are more easily seen and quantified for certain variables and/or among certain regions and/or during certain time intervals. Thus, this visual impression must be detailed and confirmed by statistical analysis: are the regions or the minutes significantly different? Are these regional differences significant all along the recording? If we consider a particular recording site, is the time evolution significant? These questions will be addressed for all retained variables.

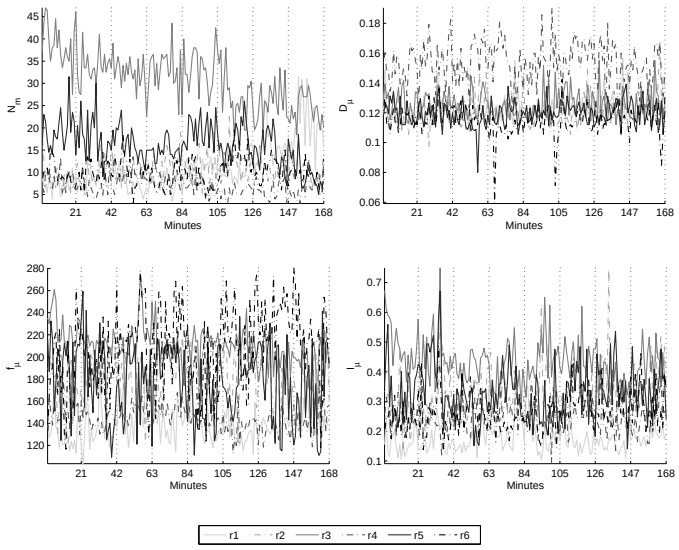


Fig. 6. Evolution of the four selected variables (N_m, D_μ, f_μ, I_μ) during 168 minutes for the 6 regions (median values over the 14 volunteers)

A very critical issue, which can lead to erroneous interpretations, is the experimental design, which must take into account the nature of the analyzed data.

First of all, as the variables (activity indices) are not Gaussian, non-parametric statistical tests must be used. Different ANOVA-like non-parametric tests have been developed and compared in the literature [36–40], and it is generally accepted that rank transforming the data (i.e., taking ranks instead of actual values) can provide robust solutions. This rank transformation can be done in different ways: either (1) globally on the whole data, (2) after aligning the data to eliminate the influence of one of the factors (i.e., subtraction of lines means before testing for column differences, for example), or (3) after ranking separately by factors (i.e., ranking the values on each line separately, instead of ranking the whole data matrix). Intuitively, rank alignment (2) applied on the third factor (patients) will normalize the volunteers by considering equal means, while the separate ranking (3), known as Friedman test, provides almost the

same effect, but normalizes both the means and the variances of the patients.

Second, the previous questions have to be translated into a statistical methodology, applied separately and successively for each variable (I_μ, N_m, f_μ and D_μ):

A. Gross analysis. Perform separate one-way Kruskal-Wallis tests (equivalent to ANOVA on not aligned ranks [36]) to assess differences among regions considering as replicates all values for a given region, regardless of the patient and minute (respectively among minutes, considering as replicates all values for a given minute, regardless of the region and patient).

B. Global analysis. The previous simple approach eludes the possible influence of the recording site on the time evolution and of the recording moment on the regional activity. In fact, a more complex design should take into account the three variable factors of the data-set: the regions, the minutes and the patients. Still, the three factors do not have the same nature: we are interested in differences among regions and among minutes, but not among patients, as our starting hypothesis was that they are all issued form the same healthy population. Nevertheless, we should take into account the inter-patient variability when testing for differences among regions and/or minutes. In this design, the third factor is what is called a random factor, and a three-way test with two fixed factors (6 regions and 168 minutes) and a random factor (14 patients) should be performed. Perform then a 3-ways ANOVA on ranks (not-aligned), with 2 fixed factors and one random, to test for differences among regions, among minutes, and for possible *interactions* between them. This last term is very critical, because if interaction exists, one should consider separate sets of data by region (respectively by minute) and make a "fine analysis" as described next.

C. Fine analysis. If interactions ar high, Zar [36] recommends to test only for differences between individual cells. This approach would be fastidious and useless: even if the difference is significant, what information could we extract by knowing that a particular minute recorded in a particular region is different from another minute recorded elsewhere? A mid-point solution is to group cells by "families" and test for differences among them: are the regions different, tested minute by minute? Are the minutes different within a particular region (i.e., is there a significant time evolution of the considered activity index for a given region)? This approach implies separate two-way ANOVA on ranks: we only consider data issued from one of the levels of a given fixed main factor and perform the test according to the other fixed main factor and the random one. For example, we could consider all regions recorded during a given minute and perform a two way analysis with one fixed factor (regions) and one random factor (patients). The previously defined 3D matrix will then split in several 2D matrices. For each of the 168 minutes, matrices having 6 lines (regions) and 14 columns (patients) will be obtained (a similar approach

leads to 168×14 matrices for each of the 6 regions).

A more robust alternative is possible: separate ranking by factors (leading to Friedman test, implemented here using the statistic given by [36] for repeated measures), leads to a reduction of the interaction, and therefore should be used if the first solution (ANOVA on ranks) reveals a high degree of interaction. Again, this analysis is made for each of the 168 minutes to test for differences among regions; a similar approach is implemented after considering separate sets of data for each region, to test for differences among minutes.

One final issue before presenting the results. The described tests allow to check for differences among *several* groups, but not between *two* particular groups: they provide the information that at least two groups are different, without indicating which. Multiple comparisons procedures [36, 37] must be used to verify this point. Some authors [36] recommend to use them only if global ANOVA-type tests indicate significant differences, while others [41] have the completely opposite opinion. Among these tests, the current approach on ranked data (after Kruskal-Wallis or Friedman) is the Nemenyi non-parametric multi-comparison [36], which we used when these tests were applied. The results of multiple comparisons are presented as suggested by [36], by underlining together similar groups (not significantly different). For example, $\underline{X_1 X_2} \underline{X_3 X_4}$ signifies that X_1 is different from X_3 and X_4 but is similar to X_2, which is similar to X_3 also, and so on. As seen in this example, a current problem when using multiple comparisons is their transitivity. We decided to include it in our interpretation of the results: if X_1 is similar to X_2 and X_2 is similar to X_3, then X_1 is similar to X_3. In our example, everything is similar, so we will rather note this as $\underline{X_1 X_2 X_3 X_4}$ and we will say that X_1, X_2, X_3 and X_4 are found similar after a transitive multiple comparison. Unless explicitly stated, all multiple comparison results in this paper are transitive. This choice might seem to restrictive, but, having in mind the actual state of knowledge on the abdominal sounds, we have decided to favor statistical validity of our findings with the risk of loosing more subtle and detailed interpretations[5].

IV. Experimental Results

Figure 6 presents the time evolution for the retained variables and for each region. Median values computed over the 14 volunteers are displayed. As one can see, the third region seems richer in sounds than the others, and especially than the first one (see Fig. 1 for positioning). This difference is present for the first part of the recording, but it attenuates at the end. Moreover, as the dispersion of the patients

[5]For example, groups $X_1 X_2$ and $X_3 X_4$ could be considered different

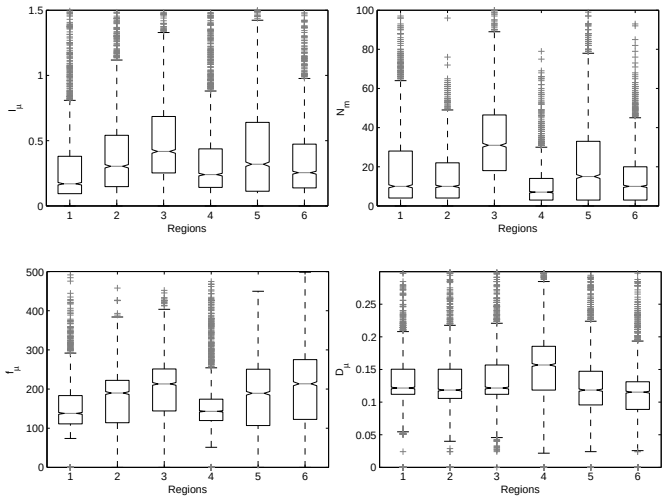


Fig. 7. Box-plots of the 4 variables (I_μ, N_m, f_μ and D_μ) for the 6 regions (all minutes included). The medians for each region are represented with their confidence intervals (notches) to facilitate multiple comparisons. Box limits are at 25 and 75 percentiles (i.e., 50% of the minutes are "inside" the box), while '+' signs represent outliers (minutes situated outside the limits defined by extending the box, in both directions, by 1.5 its size)

around median values is not displayed, we cannot visually appreciate if this difference is significant or if it is masked by a high patient inter-variability. This is the role of the statistic test, and their results are presented hereafter.

A. Gross Analysis

A possible quantification of the differences among the main factors (regions and minutes), completely ignoring any possible interactions, can be done by Kruskal-Wallis tests. As expected, minutes cannot be separated, all p values are superior to 0.05. On the contrary, the regions are significantly different for all of the four considered variables.

Still, the outputs of the KW tests indicate that a difference exists, without specifying which precise regions are different. Multiple comparisons using the Nemenyi test were then performed and the results can synthetically be presented using the underlining notation introduced in the previous section. According to the four activity indices, the regions show the following differences: for the sound intensity I_μ: $r_3\underline{r_2 r_5}\ \underline{r_6 r_4} r_1$, for the number of sounds N_m: $r_3 r_5 \underline{r_1 r_2 r_6} r_4$, for the median frequency f_μ: $r_3 r_6 \underline{r_5 r_2}\ \underline{r_4 r_1}$ and for the median sound duration D_μ: $r_4 \underline{r_3 r_1 r_2} r_5 r_6$.

Synthetically, the third region emits more sounds than the others, with higher frequencies and intensities, while the fourth region has longer sounds, but very few. The sound intensity can not be used to distinguish

between the second and fifth regions, nor the number of sounds between the first, second and sixth. Different other interpretations are left to the reader, but we remind that no interaction is taken into account here, so only very global and approximate statements can be made concerning the differences among regions and the absence of difference among sequences (see Fig. 7).

The previous multiple comparisons results confirm the visual analysis suggested by Fig. 6: if the auscultation is performed during a long enough period, the different abdominal regions have statistically distinguishable activity. All of the selected features can be used, and the regions can generally be sorted according to these features (although the second region, for example, cannot be individually separated from the others).

B. Global Analysis

A more complete approach is a 3-way ANOVA with two fixed factors (regions and minutes) and one random factor (patients), performed after rank transformation on each of the 4 variables. This step permits to evaluate the interactions: if they are not present, the differences revealed by this method would be sufficient for the phonoenterogram analysis.

For f_μ, the obtained p-values indicate very high levels of interaction among all factors, so no analysis on main factors can be performed. For I_μ and D_μ, very high interactions exist between regions and patients and between minutes and patients, but no significant value appears for regions-minutes interaction ($p = 0.26$ and $p = 0.36$ respectively). Still, both main factors have a high interaction with the patients, so interpreting main factors might be misleading for these variables also. For N_m, the results are similar, with high interactions between regions and minutes and between minutes and patients. Ignoring the interactions, it seems that all variables permit to detect significant differences among regions ($p = 0.005$, $p = 0.002$, $p = 0.0005$ and $p = 0.0007$ for I_μ, N_m, f_μ and D_μ respectively) but not among minutes ($p = 0.97$, $p = 0.99$, $p = 0.65$ and $p = 0.73$).

C. Fine Analysis

As described in the previous section, in case of interaction data are considered separately. A first option is to test for differences among regions for each minute to confirm and detail the results presented for the gross analysis (subsection IV-A). The second option is complementary: consider regions separately and test for differences among minutes.

1) Regional activity distribution during digestion: This subsection aims to find if the regions can be considered statistically different after a one minute auscultation, and if so, which minute after the meal is

the most appropriate. Unfortunately, testing for differences among regions for a given minute (Friedman test) does not give any significant result: one minute of auscultation is not sufficient to distinguish between abdominal regions, regardless of the activity index. This result seems to contradict the information from Fig. 6: it seems quite clear that, for N_m for example, the third region is higher than the others for the first part of the recording. The same observation can be made for f_μ (regions r_1 and r_4 lower than r_6 around minute 120) or for D_μ (region r_4 higher than r_6 around minute 70). On the other hand, the curves from Fig. 6 show certain natural trends, which seem to indicate that successive minutes belong to similar physiologic conditions. Therefore, we have decided to concatenate several minutes to form sequences and to test further on for differences among regions by sequence (instead of by minute). Two solutions are possible: either recompute the activity indices for the new time interval, or consider the minutes belonging to a sequence as statistical replicates (i.e., representative measures for the given sequence). We adopted the second solution: on one hand, it preserves the physiological activity indices defined in the literature and used for the previous results and, on the other hand, it improves the statistic tests reliability.

Different lengths for the tested sequences (analysis windows) have been considered, from 2 minutes to 42 minutes. As a first approach and to ease the interpretation, we did not consider sliding windows, but contiguous and disjoint (i.e., 84 to 4 different sequences). It is clear that, for long sequences, the time position (when to analyze) is less precise, as the analysis comes close to the gross Kruskal-Wallis tests. Moreover, as the activity indices are non-stationary (Fig. 6), long sequences might mask regional differences that are more important during certain digestion phases. Therefore, in our opinion, the optimal length of a sequence should be a compromise: long enough to obtain statistical significance, but as short as possible, to have a good time resolution and stationary signals. Still, as no *a priori* knowledge exists and having these considerations in mind, we present here the different obtained results, with a particular focus on sequences having 21 minutes length, considered optimal.

For sequences having a two-minute length, the results start to confirm the gross Kruskal-Wallis observations. For example, for the N_m variable, the Friedman test has a significant p-value for almost all sequences (i.e., there are differences among regions) and Nemenyi multiple comparisons indicate that the third region r_3 produces more sounds than the others. In particular, a two-minute auscultation during the first 10 minutes after the meal, as well as between minutes 60 to 80, should permit to order the regions: r_3 is the richest (and significantly different), followed by r_5.

The number of sounds by minute N_m is not the only variable permitting to distinguish among regions: D_μ indicates that the fourth region r_4 produces significantly longer sounds around minutes 60 and 120.

Three- and five-minute sequences confirm these findings. For example, for five-minute sequences, I_μ

indicates that r_3 is significantly louder than the others during the first 10 minutes and between minutes 90 to 110. The sound duration D_μ is generally longer for r_4, and this is constantly and significantly evident two hours after the meal (minutes 115 to 125).

Longer sequences reinforce the presented results. A particularly interesting case is obtained for 21 minutes length sequences, which split the recording into eight equal parts: the differences given by the statistical tests have a degree of significance similar to those obtained for the whole length. For all the sequences, the loudest region (I_μ) is r_3 (corresponding to the ileocecal valve, i.e., gut-colon junction) and the most quiet is r_1 (stomach or upper colon). Transitive multiple comparisons using Nemenyi tests allow individual separation of r_3 from all other regions for all sequences. Region r_1 (the most quiet) can also be separated according to these tests during the first hour (sequences 1 to 3). Globally, *the third region is significantly louder and the first region is significantly quieter than the others at the beginning of the digestion, and the third region remains so during all the recording period*. Again, this conclusion gives a statistic confirmation to the visual impression from Fig. 6, where these differences among regions are more or less visible during the recording.

Almost similar conclusions can be obtained for the second variable (N_m). It is still r_3 *which is the richest in sounds while the poorest is* r_4 *(lower central abdomen), except for the first sequence, when* r_1 *has the lowest number of sounds*. Multiple comparisons are more significant at the beginning and at the middle of the digestion than at the end (pre-prandial period): r_3 is significantly different from all the others from the first sequence until almost 2:30 hours after the meal (first 7 sequences), but it cannot be distinguished from r_1 (stomach) for the last 21 minutes (when most of the volunteers were hungry, with their stomach gurgling). During the last four sequences, the fourth region r_4 was also significantly poorer.

From a frequency f_μ point of view, the regions are significantly different during all sequences. *The third and the sixth region are the highest in frequency (*r_3 *during the first hour and* r_6 *during the last hour), while* r_4 *and* r_1 *are the lowest (*r_1 *during sequences 1, 2, 4-6 and* r_4 *during sequences 3, 7 and 8)*. Nevertheless, according to Nemenyi tests, r_3 is significantly higher than all the others only during the first 21 minutes ($r_3\underline{r_5 r_6}\ \underline{r_2 r_4 r_1}$) and r_6 is significantly higher only during sequence 7 ($r_6 r_3 \underline{r_5 r_2}\ \underline{r_1 r_4}$).

The analysis performed on D_μ shows that regions are significantly different according to both Friedman and ANOVA on ranks tests all along the recording, except for the last 21 minutes, when ANOVA output becomes insignificant. *The fourth region (gut) has the longest sounds during all the digestion*, and it is significantly different from all others (Nemenyi test) for the first approximately 2:30 hours after the meal (sequences 1 to 7). *The shortest sounds are generated in regions 5 and 6*, but they are significantly different

TABLE I

REGION ORDERING AND MULTIPLE COMPARISONS BY SEQUENCE (21 MINUTES LENGTH) AND BY ACTIVITY INDEX

Seq.	I_μ	N_m	f_μ	D_μ
s_1	$r_3 r_5 r_6 r_4 r_2 r_1$	$r_3 r_5 r_6 r_2 r_4 r_1$	$r_3 r_5 r_6\ r_2 r_4 r_1$	$r_4 r_3 r_1 r_5 r_2 r_6$
s_2	$r_3 r_5 r_2 r_6 r_4 r_1$	$r_3 r_5 r_6 r_2 r_1 r_4$	$r_3 r_6 r_5 r_2 r_4 r_1$	$r_4 r_3 r_1 r_2 r_6 r_5$
s_3	$r_3 r_2 r_4 r_5 r_6 r_1$	$r_3 r_5 r_2 r_6 r_1 r_4$	$r_3 r_6\ r_5 r_2 r_1 r_4$	$r_4 r_3 r_2 r_1\ r_6 r_5$
s_4	$r_3 r_5 r_2 r_4 r_6 r_1$	$r_3 r_5 r_2 r_6 r_1 r_4$	$r_6 r_3 r_5 r_2 r_4 r_1$	$r_4 r_2 r_1 r_3 r_5 r_6$
s_5	$r_3 r_5 r_2 r_4 r_6 r_1$	$r_3 r_5 r_2 r_1 r_6 r_4$	$r_3 r_6 r_2 r_5 r_4 r_1$	$r_4 r_3 r_1 r_2 r_5 r_6$
s_6	$r_3 r_5 r_6 r_2 r_4 r_1$	$r_3 r_5 r_6 r_2 r_1 r_4$	$r_6 r_3 r_5 r_2 r_4 r_1$	$r_4 r_3 r_1 r_2 r_5 r_6$
s_7	$r_3 r_2 r_6 r_5 r_4 r_1$	$r_3 r_1 r_5 r_2 r_6 r_4$	$r_6 r_3 r_5 r_2\ r_1 r_4$	$r_4 r_1 r_3 r_2 r_5 r_6$
s_8	$r_3 r_2 r_6 r_5 r_4 r_1$	$r_3 r_1\ r_5 r_2 r_6 r_4$	$r_6 r_3 r_5 r_2 r_1 r_4$	$r_4 r_3 r_2 r_1 r_5 r_6$

only at the middle of the digestion (seq. 3: $r_4 r_3 r_2 r_1\ r_6 r_5$, seq 4: $r_4 r_2 r_1 r_3 r_5 r_6$, seq 5: $r_4 r_3 r_2 r_1 r_5 r_6$).

In order to enforce the statistical validity of the results an to ease the lecture, all the previous multiple comparisons were *transitive*. The results of the non-transitive multiple comparisons are not detailed here, but a synthetic view is shown in table I, which completely presents the inter-region differences for a sequence length of 21 minutes. All multiple comparisons (for the 4 variables and the 8 sequences) are represented by underlining similar regions.

2) Activity time evolution by region: Following the same approach, we present here a detailed analysis by region, the goal being to separate among minutes (or sequences) inside each region. In other words, we attempt to propose a phonoenterogram segmentation based on statistical characteristics.

As in the previous subsection (IV-C1), a first approach considers equal length sequences and varies this length in order to find the necessary (minimal) value for a reliable statistical analysis. Keeping this duration small allows to preserve a good time resolution for a possible segmentation of the phonoenterograms in digestion phases: indeed, for each variable, *consecutive similar* time sequences could be associated, according to the results of the statistical tests, to define physiological digestion phases, while the frontiers between two phases would be placed between two *consecutive different* sequences[6]. It is clear that, as the (constant) length of the sequence increases, the resulting segmentation becomes sub-optimal, both because of the phonoenterogram "sub-sampling" and of its non-stationarity. Nevertheless, these segmentation results lead to a first approximation of the abdominal activity by piecewise constant functions: curves from Fig. 6 could be approximated by constant fixed length segments. Similar level segments (i.e., not

[6]Moreover, the shorter the necessary duration for a reliable statistical analysis, the easier the clinical implementation. Even if for the moment is speculative and further tests are needed, this approach should permit to suggest clinical guidelines for the abdominal auscultation.

significantly different according to the tests) can then be concatenated to define a longer phase, the obtained result being a first segmentation of the phonoenterogram in statistically different digestive phases.

Clearly, different other approaches can be adopted. For example, variable length sequences could be segmented directly on the curves from Fig. 6, based on some consistency criteria (trend changes, piecewise linear regression), and the resulting digestive phases could be compared among them by statistical tests to assess the validity of the proposed segmentation. A third method could propose physiologically defined phases, according to the actual medical knowledge on the digestion (phases of the migrating motor complex for example), with again an *a posteriori* statistical validation. Both these last two approaches might possibly offer a more precise segmentation of the digestion, but on the other hand, they need a consecutive statistical validation which risks to concatenate them if the necessary level of significance is not reached. In our opinion, implementing statistical tests to check for significant differences among trends (or slopes of piecewise linear regressions) exceeds the aim of this paper and the actual state of knowledge on the abdominal sound activity. Therefore, in this paper, the first described method (constant piecewise segments) is only implemented: although less precise, it offers directly the needed statistical significance.

As expected, sequences of one minute length cannot be significantly distinguished regardless of the region or of the activity index. We adopted the same approach, concatenating the minutes into longer equal size contiguous sequences and testing for differences among them. As argued in the previous paragraph, this leads to an approximate segmentation of the phonoenterogram (by variable, for each region).

The shortest sequences that can be differentiated last 15 minutes (11 sequences for the first 165 minutes), but this is only possible for one variable (N_m) and one region (r_1). More precisely, Friedman test gives a $p < 0.01$ and multiple comparisons yield the following order: $\underline{s_{11}s_{10}s_9}\ \underline{s_7s_8s_6s_3s_2s_1s_5s_4}$, which means that during the last 45 minutes (i.e., starting two hours after the meal), the region r_1 is significantly richer in sounds than during the first 2 hours.

Increasing the sequence duration improves the results of the tests. For 21 minutes sequences (8 sequences), several regions display significant time evolution, mainly concerning the number of sounds N_m. Multiple comparisons with Nemenyi test permit to individually distinguish digestion phases for two regions: for r_1, the last sequence (seq. 8) is the richest in sounds, followed by sequence 7 ($s_8s_7\underline{s_6s_5s_1s_2s_4s_3}$); for r_3, the evolution of the number of sounds follows an inverse path: the richest is the first sequence, immediately after the meal, and the poorest is the eighth, when the digestion is probably finished and the volunteers are right before their next meal ($s_1\underline{s_2s_3s_4s_5}\ \underline{s_6s_7s_8}$). It is quite remarkable that these evolutions show constant natural trends: the stomach makes more sounds when empty (i.e., when the

TABLE II

REGION ORDERING AND MULTIPLE COMPARISONS BY SEQUENCE (21 MINUTES LENGTH) AND BY ACTIVITY INDEX

Reg.	I_μ	N_m	f_μ	D_μ
r_1	$\underline{s_8 s_4 s_6 s_7 s_5 s_3 s_2 s_1}$	$s_8 s_7 \underline{s_6 s_5 s_1 s_2 s_4 s_3}$	$\underline{s_3 s_4 s_1 s_8 s_2 s_5 s_7 s_6}$	$\underline{s_7 s_8 s_6 s_5 s_4 s_3 s_2 s_1}$
r_2	$\underline{s_8 s_3 s_7 s_2 s_4 s_5 s_6 s_1}$	$\underline{s_4 s_6 s_3 s_5 s_7 s_8 s_2 s_1}$	$\underline{s_8 s_2 s_5 s_4 s_3 s_6 s_7 s_1}$	$\underline{s_8 s_7 s_4 s_3 s_2 s_6 s_5 s_1}$
r_3	$\underline{s_1 s_5 s_3 s_2 s_8 s_6 s_4 s_7}$	$s_1 \underline{s_2 s_3 s_4 s_5 s_6 s_7 s_8}$	$\underline{s_1 s_4 s_3 s_5 s_7 s_6 s_2 s_8}$	$\underline{s_8 s_7 s_4 s_3 s_2 s_6 s_5 s_1}$
r_4	$\underline{s_3 s_4 s_1 s_2 s_5 s_6 s_8 s_7}$	$\underline{s_4 s_3 s_1 s_2 s_5 s_7 s_8 s_6}$	$\underline{s_1 s_4 s_3 s_2 s_5 s_8 s_7 s_6}$	$\underline{s_2 s_6 s_3 s_4 s_5 s_1 s_7 s_8}$
r_5	$\underline{s_2 s_1 s_4 s_7 s_6 s_8 s_5 s_3}$	$\underline{s_5 s_4 s_1 s_2 s_3 s_6 s_7 s_8}$	$\underline{s_1 s_4 s_3 s_2 s_5 s_8 s_7 s_6}$	$\underline{s_1 s_2 s_6 s_5 s_4 s_7 s_8 s_3}$
r_6	$\underline{s_8 s_7 s_3 s_6 s_1 s_2 s_5 s_4}$	$\underline{s_6 s_3 s_7 s_2 s_4 s_8 s_5 s_1}$	$\underline{s_7 s_4 s_6 s_8 s_3 s_5 s_1 s_2}$	$\underline{s_2 s_8 s_7 s_3 s_6 s_1 s_5 s_4}$

person is hungry), while the lower-right abdomen (end of the gut, ileocecal valve) is very active at the beginning of the digestion at its activity decreases constantly until the end.

Finally, splitting the recording in only 4 sequences having 42 minutes length each, all variables become significant, depending on the region. For r_1, the last 42 minutes are the richer in sounds and the loudest (for N_m we have $s_4 s_3 \underline{s_1 s_2}$ while for I_μ $\underline{s_4 s_3 s_2 s_1}$), although the sound intensity is not significantly higher according to Nemenyi test, except compared with s_1 (first 42 minutes). The number of sounds permits to establish significant time evolutions also for regions 2, 3, 4 and 5 (r_2 : $\underline{s_2 s_3 s_4 s_1}$, r_3 : $s_1 s_2 s_3 s_4$, r_4 : $s_2 s_1 \underline{s_3 s_4}$, r_5 : $\underline{s_3 s_1 s_2 s_4}$). Frequency evolution can be observed for r_1 : $s_2 \underline{s_1 s_4 s_3}$, r_4 : $\underline{s_2 s_1 s_3 s_4}$ and r_6 : $\underline{s_4 s_2} s_3 s_1$ and median sound duration for r_1 : $\underline{s_4 s_3 s_2} s_1$ and r_4 : $\underline{s_3 s_2 s_1} s_4$. Although not always obvious, some natural trends can be detected using these long-duration sequences. The number of sounds N_m seems a valid indicator of the normal digestion evolution for almost all regions: it constantly increases for the stomach (r_1), it constantly decreases for the ileocecal region (r_3) and it has a similar evolution for the preceding segment (lower abdomen r_4). The frequency is higher during the first half of the digestion both for r_1 and for r_4. This last observation is confirmed also by a gross analysis: Kruskal-Wallis and Nemenyi tests for sequence differences, regardless of the region and of the patient, indicate that s_2 (second quarter of the recording) has significantly higher frequencies than the other periods: $s_2 \underline{s_1 s_4 s_3}$.

Interestingly enough, the segmentations resulting from different sequence lengths confirm each other in most of the cases, or at least they are complementary: for example, taking the first region r_1 and variable N_m, all considered sequences (15, 21 or 42 minutes) lead to the conclusion that the last part of the phonoenterogram (40 to 45 minutes) is statistically different from the previous sequences, with complementary information given by the last two analysis (21 and 42 minutes).

A synthetic presentation of the inter-sequence differences, including all the multiple comparisons results is presented table II for 21 minute-length sequences.

V. Conclusion and Future Research

This article addresses two main issues: the abdominal sounds processing methodology and their detailed statistical analysis.

Several methodological signal processing steps were proposed and/or improved and the resulting chain was employed to extract physiologically meaningful data from the long-term multi-channel phonoenterograms. The proposed solution is effective, although an interesting perspective research could be the comparison of the different other signal processing methods developed in the recent literature for abdominal sound analysis. However, the processing steps proposed in this paper proved to be reliable enough to furnish statistically consistent information, analyzed in the second part of this work.

Two types of analysis were performed, aiming to check if and under which conditions abdominal auscultation can furnish statistically reliable data on the normal digestion, both in terms of localization and in terms of time evolution.

According to our results, abdominal activity in different regions can be distinguished using abdominal auscultation. It seems that, to differentiate between regions, rather short-term auscultation (3 to 5 minutes) can be sufficient, especially (but not only) when this auscultation is performed immediately after the meal. For example, for normal digestion, the third considered region r_3 (lower right abdomen, ileocecal valve) should be louder, should emit more sounds and have higher frequencies. Approximately two hours after the meal (digestion final phases), the sounds emitted in the lower central abdomen r_4 are significantly longer than those from the other regions. Of course, future validation on healthy and pathologic cases is needed, but present results indicate that these findings consistently describe normal functioning of the abdominal tract. A comparative view of all results suggests some interesting information for clinical auscultation: it seems that the most informative regions, at least for analyzing normal digestion, are r_1, r_3 and r_4 (stomach, ileocecal region and gut region); the most effective activity index is the number of sounds N_m, although complementary information is carried by the other indices; finally, better results can be obtained by an immediately post-prandial auscultation, although later phases are also informative, especially if the auscultation is longer.

A more difficult issue is the analysis of the digestion evolution over time using the selected physiological activity indices. Although some trends can be detected and thus individual minutes or short sequences of auscultation seem different at the beginning and at the end of the digestion process (especially when counting the sounds emitted by the stomach and by the lower right abdomen), we were not able to prove statistically this difference. Even if digestion evolution cannot be clearly evaluated by realistic (short-

time) clinical auscultation, long-time phonoenterograms can do it. Indeed, considering longer auscultation sequences (unrealistic in clinical environment but possible using an automatic system), these trends can be detected from the recorded data and they become significant, not only considering the number of emitted sounds but also in duration, frequency and intensity. This point also needs further validation, and we are confident that a more extensive data-base can improve the statistical reliability of our results.

An important point not completely addressed in the current work is the length of the auscultation sequence needed to distinguish among regions or among sequences themselves. Our current proposal was to consider equal length sequences, which is a coherent approach from a statistical point of view: estimation made over populations having similar sizes have similar properties (confidence intervals), and further comparisons are facilitated. Nevertheless, a physiologically justified approach would consist in a previous segmentation of the time evolution of the activity indices, in order to detect possible natural digestion phases, which can be further on tested for statistical differences using a more elaborate test methodology. Hopefully, this approach should permit to propose and confirm a finer segmentation of the digestion phases and will be addressed in a future work.

An important future research direction is the validation of the proposed signal processing and data analysis methodology on a more extensive data-base and on pathological cases, leading to a possible increase of the statistical validity. Still, the results presented in this paper convey statistical evidence that the regional abdominal activity of healthy patients shows a certain structure. More precisely, a time-evolving activity pattern can be established for healthy patients, and one can speculate that changes in this pattern reflect a modified patient state (health, digestion phase, digestion habits).

For the moment, no pathological case was recorded using the standardized protocol, but a short phonoenterogram of a patient with gastritis was easily distinguished from the others (higher sound intensity and number, especially in r_1). As our method allows a precise analysis of the normal abdominal functioning, we are confident that pathologic recordings can also be characterized. An important problem is the creation of a standard healthy phonoenterogram data-base, necessary for any further comparative study.

In our opinion, the main conclusion of this work is that normal (and possibly pathologic) gastrointestinal activity might be analyzed using abdominal auscultation, but this should be done with great care: if the region differences can be assessed by rather short-time auscultation, and thus in a clinical environment, digestion evolution evaluation needs longer recording periods and thus an automatic tool.

REFERENCES

[1] W. Cannon, "Auscultation of the rythmic sounds produced by the stomach and intestines," *American Journal of Physiology*, vol. 14, pp. 339–353, 1905.

[2] J. Farrar and F. Ingelfinger, "Gastrointestinal motility as revealed by study of abdominal sounds (with discussion)," *Gastroenterology*, vol. 29, no. 5, pp. 789–802, 1955.

[3] C. Garner and H. Ehrenreich, "Non invasive topographic analysis of intestinal activity in man on the basis of acoustic phenomena," *Research in Experimental Medicine*, vol. 189, no. 2, pp. 129–140, 1989.

[4] H. Yoshino, Y. Abe, T. Yoshino, and K. Ohsato, "Clinical application of spectral analysis of bowel sounds in intestinal obstruction," *Dis. Col. Rectum*, vol. 33, no. 9, pp. 753–757, 1990.

[5] T. Tomomasa, A. Morikawa, R. Sandler, H. Mansy, H. Koneko, T. Masahiko, P. Hyman, and Z. Itoh, "Gastrointestinal sounds and migrating motor complex in fasted humans," *American Journal of Gastroenterology*, vol. 94, no. 2, pp. 374–381, 1999.

[6] B. Craine, M. Silpa, and C. O'Toole, "Computerized auscultation applied to irritable bowel syndrome," *Digestive Diseases and Sciences*, vol. 44, no. 9, pp. 1887–1892, 1999.

[7] ——, "Two-dimensional positional mapping of gastrointestinal sounds in control and functional bowel syndrome patients," *Digestive Diseases and Sciences*, vol. 47, no. 6, pp. 1290–1296, 2002.

[8] C. Liatsos, L. Hadjileontiadis, C. Mavrogiannis, D. Patch, S. Panas, and A. Burroughs, "Bowel sounds analysis: a novel noninvasive method for diagnosis of small-volume ascites," *Digestive Diseases and Sciences*, vol. 48, no. 8, pp. 1630–1636, 2003.

[9] D. Du Plessis, "Clinical observation on intestinal motility," *South African Medical Journal*, vol. 28, pp. 27–33, 1954.

[10] C. Dimoulas, G. Kalliris, G. Papanikolaou, V. Petridis, and A. Kalampakas, "Bowel-sound pattern analysis using wavelets and neural networks with application to long-term, unspervised gastrointestinal motility monitoring," *Expert Systems with Applications*, vol. 34, pp. 26–41, 2008.

[11] E. Arnbjörnsson, "Normal and pathological bowel sound patterns," *Ann. Chir. Gynaecol.*, vol. 75, no. 6, pp. 314–318, 1986.

[12] R. Ranta, V. Louis-Dorr, C. Heinrich, D. Wolf, and F. Guillemin, "Principal component analysis and interpretation of bowel sounds," in *26th Ann. Int. Conf. of the IEEE-EMBS*, San Francisco, USA, Sep. 2004.

[13] ——, "A complete toolbox for abdominal sounds signal processing and analysis," in *3rd European*

Medical and Biological Engineering Conference IFMBE-EMBEC'05, Prague, Czech Republic, Nov. 2005.

[14] J. Gade, P. Kruse, O. Trier Andersen, S. Boel Pedersen, and S. Boesby, "Physicians abdominal auscultation – a multi-rater agreement study," *Scandinavian Journal of Gastroenterology*, vol. 33, no. 7, pp. 773–777, 1998.

[15] M. Yuki, K. Adachi, H. Fujishiro, Y. Uchida, Y. Miyaoka, N. Yoshino, T. Yuki, M. Ono, and Y. Kinoshita, "Is a computerized bowel sound auscultation useful for the detection of increased bowel motility?" *American Journal of Gastroenterology*, vol. 97, no. 7, pp. 1846–1848, 2002.

[16] R. Ranta, C. Heinrich, V. Louis-Dorr, and D. Wolf, "Interpretation and improvement of an iterative wavelet–based denoising method," *IEEE Signal Process. Lett.*, vol. 10, no. 8, pp. 239–241, 2003.

[17] R. Ranta, V. Louis-Dorr, C. Heinrich, D. Wolf, and F. Guillemin, "Towards an acoustic map of abdominal activity," in *25th Ann. Int. Conf. of the IEEE-EMBS*, Cancun, Mexico, Sep. 2003.

[18] R. Ranta, V. Louis-Dorr, C. Heinrich, and D. Wolf, "Iterative wavelet–based denoising methods and robust outlier detection," *IEEE Signal Process. Lett.*, vol. 12, no. 8, pp. 557–560, 2005.

[19] C. Dimoulas, G. Kalliris, G. Papanikolaou, and A. Kalampakas, "Long-term signal detection, segmentation and summarization using wavelets and fractal dimension: A bioacoustics application in gastrointestinal-motility monitoring," *Computers in Biology and Medicine*, vol. 37, pp. 438–462, 2007.

[20] S. Kraman, G. Wodicka, G. Pressler, and H. Pasterkamp, "Comparison of lung sound transducers using a bioacoustic transducer testing system," *Journal of Applied Physiology*, vol. 101, pp. 469–476, 2006.

[21] H. Mansy and R. Sandler, "Detection and analysis of gastrointestinal sounds in normal and small bowel obstructed rats," *Medical and Biological Engineering and Computing*, vol. 38, no. 1, pp. 42–48, 2000.

[22] L. Hadjileontiadis and S. Panas, "Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 12, pp. 1269–1281, 1997.

[23] L. Hadjileontiadis, L. Liatsos, C. Mavrogiannis, T. Rokkas, and S. Panas, "Enhancement of bowel sounds by wavelet-based filtering," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 876–886, 2000.

[24] R. Ranta, C. Heinrich, V. Louis-Dorr, D. Wolf, and F. Guillemin, "Wavelet–based bowel sounds denoising, segmentation and characterization," in *23rd Ann. Int. Conf. of the IEEE-EMBS*, Istambul, Turkey, Oct. 2001.

[25] L. Hadjileontiadis, "Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding–Part I: methodology," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 6, pp. 1143–1148, 2005.

[26] C. Dimoulas, G. Kalliris, G. Papanikolaou, and A. Kalampakas, "Novel wavelet domain Wiener filtering de-noising techniques: Application to bowel sounds captured by means of abdominal surface vibrations," *Biomedical Signal Processing and Control*, vol. 1, pp. 177–218, 2006.

[27] R. Coifman and M. Wickerhauser, "Adapted waveform de-noising for medical signals and images," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 5, pp. 578–586, 1995.

[28] L. Hadjileontiadis, "Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding–Part II: application results," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 6, pp. 1050–1064, 2005.

[29] S. Mallat, *A wavelet tour of signal processing*. Academic Press, 1999.

[30] ——, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[31] K. Kokkinakis and A. Nandi, "Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling," *Signal Processing*, vol. 85, pp. 1852–1858, 2005.

[32] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[33] M. Sugrue and M. Redfern, "Computerized phonoenterography: the clinical evaluation of a new system," *Journal of Clinical Gastroenterology*, vol. 18, no. 2, pp. 139–144, 1994.

[34] D. Bray, R. Reilly, L. Haskin, and B. McCormack, "Assessing motility through abdominal sound monitoring," in *19th Ann. Int. Conf. of the IEEE EMBS*. Chicago, USA, Oct. 1997, pp. 2398–2400.

[35] B. Craine, M. Silpa, and C. O'Toole, "Enterotachogram analysis to distinguish irritable bowel syndrome from Crohn's disease," *Digestive Diseases and Sciences*, vol. 46, no. 9, pp. 1974–1979, 2001.

[36] J. Zar, *Biostatistical analysis*, 3rd ed. Prentice Hall, New Jersey, USA, 1996.

[37] R. Sokal and F. Rohlf, *Biometry: the principles and practice of statistics in biological research*, 3rd ed. W.H. Freeman, New York, USA, 2001.

[38] A. Tamhane and A. Hayter, "Comparing variances of several measurement methods using a randomized block design with repeat measurements: a case study," in *Advances in Ranking and Selection, Multiple Comparisons and Reliability*. Birkhauser, 2005, pp. 165–178.

[39] H. Mansouri and G.-H. Chang, "A comparative study of some rank tests for interaction," *Compu-*

tational Statistics & Data Analysis, vol. 19, pp. 85–96, 1995.

[40] W. Conover and R. Iman, "Rank transformations as a bridge between parametric and nonparametric statistics," *The American Statistician*, vol. 35, no. 3, pp. 124–129, 1981.

[41] G. Hancock and A. Klockars, "The quest for alpha: Developments in multiple comparison procedures in the quarter century since Games (1971)," *Review of Educational Research*, vol. 66, no. 3, pp. 269–306, 1996.