ACOUSTIC LOCALIZATION OF TACTILE INTERACTIONS FOR THE DEVELOPMENT OF NOVEL TANGIBLE INTERFACES

Pietro Polotti, Manuel Sampietro, Augusto Sarti, Stefano Tubaro

Image and Sound Processing Group (ISPG)

DEI- Politecnico di Milano, Italy
`surname@elet.polimi.it`

Alain Crevoisier

Centre for Engineering and Technology Transfer

Haute Ecole Vaudoise, Yverdon, Switzerland
`alain.crevoisier@eivd.ch`

ABSTRACT.

In this paper we propose different acoustic array processing methods for the localization of tactile interactions with planar surfaces. The aim is to create a new class of tangible interfaces for musical performance that can be obtained by simply applying sensors on existing surfaces. The solutions considered in this paper are mainly based on the measurement and the analysis of the Time-Delay-Of-Arrival (TDOA) over a set of contact sensors, placed around the area of potential contact, and allows us to rapidly localize tactile interactions with reasonable accuracy.

1. INTRODUCTION

In the past two decades, research on interfaces for Human Computer Interaction (HCI) has focused mainly on audio and video analysis aimed at making the interaction with machines as natural as possible. More recently the attention has turned toward multimodal interfaces, where signals of various nature are jointly and synergistically used to convey as much information as possible between humans and machines in a bidirectional fashion. One class of signals that have not been considered as much as audio and video, are those produced by the tactile interaction with objects. The interest in such signals is, however, very strong, as they carry a formidable amount of information and could enable novel forms of expressiveness [1]. This particular area of research is addressed in depth by the European project IST-507882 TAI-CHI (Tangible Acoustic Interfaces for Computer-Human Interaction) [2]. What we present in this paper are the initial results of the TAI-CHI project concerning the detection and classification of tactile interactions with tangible interfaces. The final goal of the project is to gain a wide and versatile “vocabulary” of interactions modes involving gesture analysis and interpretation. In this sense it is not casual that the TAI-CHI project was inspired by the employment of tangible acoustic interfaces in the context of electronic music and artistic performance [1].

Tactile interactions involves the generation of vibrational patterns, depending on the interaction modality. Such vibrations take place on the object surfaces and the corresponding signals are acquired through contact sensors (typically piezoelectric sensors) applied to the surface itself. The variety of acoustic waves produced by the possible interactions provides the user with a potentially very wide, articulated and versatile vocabulary of signals that can be used for control purposes.

Different methods for the detection of interactions can be considered. These methods can be grouped into two main families, active and passive. The former is based on the evaluation of the acoustic energy that is absorbed at the points of contact, when the object is excited with ultrasound. These techniques are regarded as in-solid acoustic Holography. The latter relies on the analysis of the acoustic vibrations generated at the points of contact, when tapping or moving a finger or some tool on the surface of an object. Within this approach many techniques of source detection, localization and characterization can be developed. One example is given by the Time Reversal Process [3], based on the comparison and identification of the interaction signal with respect to a pre-recorded set of interaction signals. Another possibility is given by the Time Delay Of Arrival (TDOA) estimation, based on the analysis and comparison of signals detected by a number of sensors arranged in array or in some other convenient configuration [4], [5], [6]. In this work we consider mainly TDOA-related methods.

In the following sections we briefly describe the idea of Tangible Acoustic Interface as introduced by the TAI-CHI project (Section 2), we review TDOA localization methods and their limitations (Section 3) and we present the state of the art of our research on acoustic source localization (Section 4). Section 5 goes through the experimental results, while Section 6 deals with the lines of future research and developments. Finally in Section 7 we draw our conclusions.

2. TANGIBLE ACOUSTIC INTERFACE AS NOVEL MUSICAL INTERFACES

Tangible Acoustic Interfaces (TAI’s) lead to many possible application scenarios. As already mentioned, they were conceived as new electronic musical interfaces [1]. The main goal of using a TAI for musical applications is to re-unify the role of the generator of acoustic energy produced by the mechanical movement of the performer and that of the input interface, i.e. the controller of the sound output. In fact, in the electronic music these two roles are split: The interface (any external control device) and the sound generator (the computer) are not the same object. In other words, the ideal goal is to use the acoustic waves generated by the interaction with an object as a sound source and to employ, at the same time, the same acoustic vibrations to generate control information for processing the sound by means of a computer. Of course, this would require not only a localization of the interaction but also a refined characterization and interpretation of the interaction itself, based on pattern recognition techniques and gesture interpretation. Thanks to this, one could aim to recover a clear perceptual mapping similar to that of traditional instruments, i.e. to recover a clear correspondence between a physical gesture and the sound response produced by it. In a wider perspective, this approach will allow a formidable expansion of the gesture scope and an unlimited freedom in the definition of movements that can communicate expressive details, without the constraints of traditional instruments. In this way, it would be possible to conceive an interface according to a desired human gesture instead of adapting the gesture to a given interface. In this case the goal becomes: Matching as much as possible the naturalness of human gestures and/or the limitation of human gestures (for example, in case of disabled people).

On the other side, flexible TAI implementation techniques would allow a new freedom in the design of musical interfaces. Any form, shape object of any dimension can become an interface. In this sense, also ergonomic issues, in terms of what we could call tactile design can play a relevant role. Anyway, TAI's are not limited to musical interfaces. The highest ambition of the TAI-CHI project [2] is to bring the sense of touch to the computer world in the widest possible way. The idea is to transform physical objects, flat or complex surfaces and walls into natural and seamless touch interface by making them sensitive to tactile interactions as the skin of a human being is. After computer vision and speech/sound recognition, tactile interaction detection and classification could play a fundamental role in the design of interactive systems.

3. TDOA LOCALIZATION METHODS AND IN-SOLID WAVE PROPAGATION

TDOA-based locators are all based on a two-step procedure applied on a set of spatially separated microphones. Time delay estimation of the source signals is first performed on pairs of distant sensors. This information is then used for constructing hyperbolic curves that describe for each couple of sensors (the foci of the hyperbola) the location of all points that correspond to the estimated delay. The curves drawn for the different pairs of sensors are then intersected in order to identify the source locations. This constitutes the very simple abstract and geometrical approach to the problem. At this level sensor positions has to be selected in some clever way, not necessarily in array configuration, in order to optimize the localization resolution.

Then a number of physical phenomena arise and has to be considered in order to make the method reliable. Obviously, the performance of TDOA-based solutions depends very critically on the accuracy and the robustness of the time delay estimation (TDE). One can identify three major problems for TDOA methods for the in-solid case: background noise, number of relevant reflections (multiple sound propagation paths and, especially, dispersion.

The most crucial of these problems is the third one, i.e. the phase dispersion occurring in in-solid wave propagation. Generally speaking, waves in solid plates propagate in different ways: longitudinal and transversal waves, denoted as BAW (Bulk Acoustic Waves) and Reyleigh waves, denoted also as SAW (Surface Acoustic Waves). In thin plates one finds also other kinds of SAW: Love and Lamb waves (see for instance [7], [8]). The direction of oscillation of the particles in each one of these families of waves as well as the velocities of propagation are different. Also, each one of these types of waves can be excited according to different modes and each mode has its own dispersion curve. Thus, according to the physical interacting point on the surface and the frequency content of the excitation, different modes can be excited with significantly different propagation velocity. In other words we do not have a constant phase (or group) velocity throughout the solid surface. In solid acoustics is, thus, an extremely complex phenomenon.

In the next section we discuss the solutions we adopted in order to cope with these problems.

4. SYSTEM MODEL AND LOCALIZATION METHODS

Up to now, in order to limit the cases of the possible experimental/real-life scenarios, we considered homogeneous and isotropous materials and flat (rectangular) geometries. Also, in order to make the detection of TDOA's an easier task, we considered various kinds of impulsive excitations characterized by a sharp attack transient. The excitation signals were produced by means of screw-drivers, pencils, finger-tapping, knuckles and nail ticking. In these conditions, the simplest approach is to detect the time of the first arrival at each sensor of a signal. One has to take into account the already mentioned dispersion problems and the presence of different wave typologies, traveling at different propagation velocities. In some cases as the plexiglas, glass and metal, setting a threshold just above the background noise, is accurate enough to detect without ambiguity the arrival of the surface waves produced by an excitation. Obviously, this solution bypasses any problem due to the dispersion and loss of coherence of the wave-shape at the different sensors.

More into detail, our algorithm consists in a learning phase and an executive phase. In the first phase one has to tap on each single zone and the algorithm stores the delay values corresponding to that position. The position is then identified by a "codeword" formed by the N-1 delays, where N is the number of sensors at disposal. In the executive phase, when an interaction is detected and the corresponding set of delays of arrivals at the sensors matches a codeword within a certain tolerance, then the position of the interaction is localized as the position corresponding to that codeword. In order to avoid detections of fake arrivals, the algorithm also performs a test on the shape of the signal, which checks the sign of the derivative for a certain number (depending on the material) of samples after the candidate kick. If all of the derivative values have the same sign and are either all increasing or decreasing, the kick is passed. Also, a test, which checks the continuity of the strong signal at the kick is performed. The standard deviation of a certain number of samples after a candidate kick is compared with that of the same number of samples before the candidate kick itself. The kick is passed if the first standard deviation is greater than the latter.

In order to cope with the dispersion problems, another possible method is that of measuring the amount of (cumulating) incoming energy of the signal detected at each sensor. We already adopted this approach in a previous work [9] in the case of an amorphous wooden tables. By taking the second derivative of the cumulating energy at each sensor, we are able to characterize the time-of-arrival at each sensor as the peak of what we could think of as the "acceleration" of the incoming energy.

A completely different approach to the localization problem that exploits (instead of trying to avoid) the problem of the dispersion was also considered. The method considers the path between the interaction point and the sensor as a channel. During the learning phase, by estimating and recording the transfer function of each one of these channels, corresponding to all of the considered positions, one obtains a collection of identifier of the positions themselves. In the executive phase, the algorithm deconvolves an incoming signal by means of the pre-stored transfer functions. Only the deconvolution by means of the transfer function corresponding to the actual interaction position will give as a result an impulse-like signal (see Figure 1, the excitation signal and Figure 2a, the signal received at a sensor and deconvolved by means of the transfer function corresponding to the excitation position. All of the others will give back a dispersed signal of the kind of that of Figure 2b. In order to localize the position of the excitation, it is necessary to perform a test on the sharpness of the attack of all of the deconvolved signals in order to detect the right one.

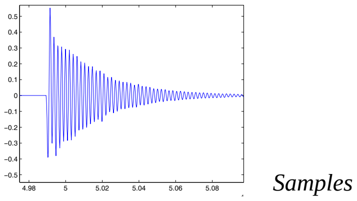

Figure 1: *Excitation signal on a plexiglas table. The excitation tool was a screw driver.*

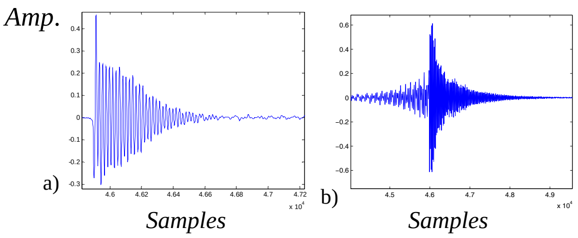

Figure 2: *a) The result of the deconvolution of a signal received at a sensor by means of the transfer function corresponding to the excitation position. The signal presents a sharp attack, similar to that of the excitation in Figure 1. b) The result of the deconvolution by means of a transfer function not corresponding to the excitation position. The signal is dispersed.*

We also studied the case of continuous and localized interactions, i.e. rubbing or scratching on a limited (less than 2x2 cm^2) area of a solid surface, for example, drawing little circles on the spot. In this case, we perform a beamforming-like test on the signals. In other words, we find the peaks of the cross correlation between each couple of sensors. The procedure is divided into a learning and executive phase as in the impulsive case. An excitation of about 1-2 seconds is applied to each position. The algorithm detects the segments of the incoming signals, having significant energy (see Figure 3). These segments are the signals, on which the algorithm performs a SB-like test or cross-correlation test. The maxima of the cross-correlation functions of the segments are calculated. A double averaging procedure is then performed over 10 segments. After the first averaging only the 50% of the lags (the one closer to the average lag) are preserved.

Finally a second averaging over the final selection of lags gives the set of data that identifies the position. These are the data that form the codewords for the continuous case. In the executive phase, when an excitation signal arrives at the contact microphones a number of segments having significant energy is considered in order to evaluate the corresponding maximal cross-correlation lags and perform the matching test for the localization of the interaction position, in the same way as in the case of impulsive interactions.

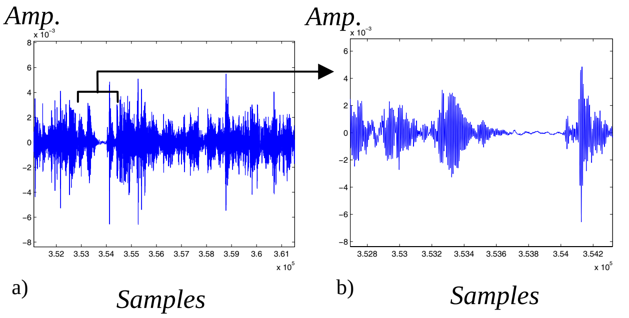

Figure 3: *a) An example of continuous excitation b) a detail from a), showing a segment where the signal is missing. The length of the interruption is about 400 samples, which could be misleading in the measurement procedure. A test on the energy of the signal allows to discard these segments, when estimating the delays.*

5. EXPERIMENTAL SETUP AND RESULTS

At present we limited our investigation to the case of flat surfaces. We tested different kinds of materials, glass, plexiglas and MDF (Medium Density Fiber), i.e. isotropous materials. We considered tables of various sizes from 130x100 cm2 to 50x50 cm^2, homogeneous with no discontinuities.

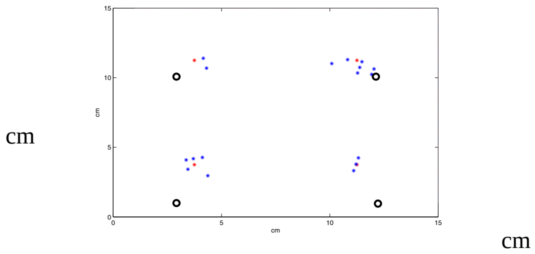

Figure 4: *Localization of 4 zones on a portion of size 15x15 cm^2 of a plexiglas table of size 50x50 cm^2. The circles represent the interaction positions, while the dots are the estimated values. The localization is performed by means of 4 sensors positioned at the corners of the square area. The maximum error was ± 1 cm.*

The measurable criteria of the proposed technical approaches are in terms of spatial resolution of localization. Currently, we have obtained a resolution of +/- 1 cm on a surface of about 15x15 cm^2 (see Figure 4).

The experimental set-up consists in an acquisition system and a set of contact sensors. The acquisition system is formed by a PC

Pentium IV 2.1 GHz, a Multi-Track hard-disk recorder a Terratec EWSMT88 sound card 8IN/8OUT working at a sampling rate of 96 KHz. New tests at a higher sampling rate of 512 KHz by means of an ultrasound acquisition card are planned for the next future. For what concerns the sensors, at present we used both a set of high quality sensors (Knowles BU-1770) and a set of 4 low-cost piezo-electric sensors. The low cost sensors provide in these early experiments a way for validation and testing the robustness of our solutions also in the perspective of possible market exploitations of the results of this research. In Figure 5 we report an example of measurements.

For the case of continuous interaction, the procedure is the following: The signals are band-pass pre-filtered according to the working band of the sensors. Then a narrower band is selected, corresponding to the frequency range with highest energy. The window employed for the energy evaluation, i.e. for the identification of a “good” long window (the window for the estimation of the cross-correlation was 10 time shorter the long one. Thus, the acceptance criterion for a long window was that 10 consecutive short windows had to pass successfully the energy threshold. At present the resolution that we considered is quite loose: ± 3.5 cm. Anyway, improvements look like to be easily achievable.

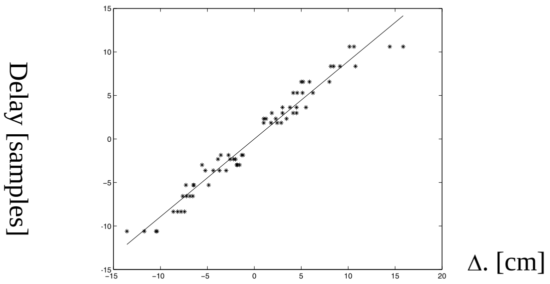

Figure 5: *Linear regression of the delay data obtained from the measurements of 16 zones, represented as a function of Δ, the distance difference between couples of sensors. The angular coefficient is 0.8660, corresponding to a mean phase velocity of 831.3 m/s.*

6. FUTURE DEVELOPMENTS

A fundamental step further of this work will be the development of a method for tracking continuous and non-localized (moving) interactions. This would extend dramatically the variety of human gestures that could be included in the tangible interaction vocabulary.

In addition to these technical improvements of the localization techniques, our research, as well as the whole TAI-CHI project, aims at the development of semantically rich tangible interfaces, by means of a low cost technique. This will require the integration of pattern recognition techniques in order to define various algorithms for the identification of the different kinds of interactions. Providing TAI-based interaction systems with the capability of discriminating between different gestures (for example, rubbing vs. scratching vs. caressing) would allow to provide with semantic and also emotional content the different kinds of interactions.

Another fundamental aspect, characterizing the meaning content and expressiveness of a gesture especially in the case of continuous interactions, are the velocity and the acceleration. These parameters will have to be estimated and interpreted as well, in order to enrich the gesture analysis.

Also, one of the TAI-CHI project goal is the implementation of all of the developed techniques on a particular software platform EyesWeb [10], developed by one of the TAI-CHI partners in the context of another European project, i.e. MEGA [11]. This will allow the algorithm to run in Real-Time and, more important, to integrate in a multimodal fashion all of the developed techniques for tactile interaction localization and characterization with the gesture analysis and recognition algorithms that EyesWeb already implements both for video and audio.

The final and ideal goal of the collection and integration of all of these techniques for the localization and classification of tactile interactions is that of bringing the expressiveness of human tactile actions and gestures in the domain of computer music.

7. CONCLUSIONS

In this paper we presented the initial results of a research project aiming at transforming ordinary objects into versatile and semantically rich control interfaces by means of an extremely low cost technique. The main idea is to analyze the acoustic signals produced by the physical interaction with (potentially) any object, whatever shape and whatever material, and exploit them as a source of information for the localization and characterization of the interaction itself. In this way, any tactile interaction with its expressive content could become a control signal for a machine. Different techniques of localization of in-solid wave sources were considered and analyzed.

8. REFERENCES

[1] Crevoisier A., Movement and Music, from traditional instruments to hyper-instruments, dissertation of the Master degree in Digital Art at the University Pompeu Fabra, Barcelona, Spain, (2001).

[2] http://www.mec.cf.ac.uk/research/pubs/taichi.html

[3] Ing R. K., Catheline S., Quieffin N. and Fink M.: Dynamic focusing using a unique transducer and time reversal process. Proceeding of the 8th International Congress on Sound and Vibration, 119-126 (2001)

[4] Haykin S., Ed., “Array Signal Processing”. Prentice Hall, (1984).

[5] Brandstein M., Ward D., Eds. “Microphone arrays – Signal processing techniques and applications”. Springer-Verlag, (2001).

[6] Stoica P., Moses R.: Introduction to Spectral Analysis. Prentice Hall (1997).

[7] Schmerr L.W.: Fundamentals of Ultrasonic Nondestructive Evaluation: A modeling Approach. Perseus Publishing, (1998).

[8] Ristic V. M., “Principles of Acoustic Devices”. John Wiley and Sons, 1983.

[9] Crevoisier A., Polotti P.: A New Musical Interface Using Acoustic Tap Tracking, MOSART Workshop on Current Research Directions in Computer Music. Barcelona, (Nov 15-16-17, 2001). http://www.iua.upf.es/mtg/mosart/papers/p26.pdf

[10] http://www.eyesweb.org

[11] http://www.megaproject.org/