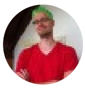
ResearchGate

Sound-based proximity detection with mobile phones

CONFERENCE PAPER · NOVEMBER 2012

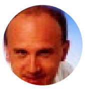
DOI: 10.1145/2389148.2389152

3 AUTHORS:



Benjamin Thiel
Universität Passau

Kamil Kloch
Deutsches Forschungszentrum für Künstlic…



Paul Lukowicz
Deutsches Forschungszentrum für Künstlic…

Sound-based Proximity Detection With Mobile Phones

Benjamin Thiel, Kamil Kloch
Embedded Systems Lab, University of Passau
Email: ben.thiel23@gmail.com, kamil.kloch@gmail.com

Paul Lukowicz
German Research Center for Artificial Intelligence (DFKI)
Email: Paul.Lukowicz@dfki.de

***Abstract*—We present a method for proximity detection with mobile phones that is based on a combination of Bluetooth communication (for the detection of coarse proximity) and sound beacons in an inaudible spectrum around 18kHz for a finer spatial resolution. The system performs a real-time recognition of personal encounters in two common situations: standing together and walking by each other. We evaluate our approach in a variety of settings ranging from office corridor, through a busy street to a shopping mall.**

I. Introduction

Knowing when people are close together is a relevant piece of information for a variety of applications ranging from social interactions recognition [1] through crowd density analysis to support of indoor localization [2]. In our work we aim to develop a system that can detect proximity with nothing more than a mobile phone in a pocket and without the need for any external infrastructure or additional sensors. While many methods have existed for proximity detection in general, there are only two possibilities that fulfill the above criteria. One is ambient sound analysis as proposed by [3]. This has the disadvantage that it requires appropriate sound to be present, which may not always be the case. Another is Bluetooth which has been extensively used for indoor localization. While special purpose hardware with elaborate facilities for the control and analysis of signal strength can provide proximity information with an accuracy of between 5 and 10 meters, standard mobile phone APIs (e.g. on the iPhone) perform significantly worse.

Experiments revealed that Bluetooth connections initiated by the iPhone API had a surprisingly high operational range, making it unsuitable for proximity detection. Two persons carrying an iPhone 4 in their pants pocket and walking around in an office building, resulted in a visibility of up to 60 meters, even through several walls. With the persons facing each other in a shopping mall, the connection was operational within a range of up to 80 meters, the longest possible distance to be found in that mall. A similar test outdoors where the two persons' backs were facing each other while approaching, an operational connection could be established at a distance of 25 meters, still a long distance considering the fact that the radio waves were shielded by two people.

In this paper we describe an algorithm which improves upon the resolution of the Bluetooth-based proximity detection method. Our solution

i) is implemented as a standalone iPhone application,
ii) is robust enough to be run on-line on the device,
iii) is reliable (99.95% True Negative rate).

The method consists of two steps. First, Bluetooth paring is used to confirm that the two devices are broadly in the same space i.e., are 'eligible' for the close-proximity test. Next, the one device starts emitting repeating sound patterns in a predefined inaudible narrow spectrum and the other tries to detect them. The method works well in a variety of environments ranging from empty office space to a busy mall.

II. Related Work

There are several infrastructure-based indoor-localisation systems using radio signals (i.e. WiFi, Bluetooth), but they are as such not relevant to the paper, since our approach is entirely infrastructureless. Localisation using sound, on the other hand, has been proposed for several use cases, though none of them tries to capture interpersonal encounters of persons walking around using only software and off-the-shelf, ubiquitous hardware like cell phones.

Scott et al. [4] propose a system where microphones are pre-installed in a room and the system detects and locates user-genareted sounds (i.e. clapping). In [3], Wirz et al. used fingerprinting to verify the existence of a relation between the distance of two devices and the similarity of the recorded ambient sound. Girod et al. present ENSBox, a custom-built platform for rapidly deploying self-calibrating distributed localisation using acoustics in [5].

Arentz et al. [6] demonstrate that near-ultrasonic sound processing is feasible on iPhones and can even be used for data transfer in the short range. In [7], Peng et al. implemented an acoustic ranging system using smartphones yielding a high accuracy (up to 2cm in the 10 meter range), though the phones were in line of sight. Acoustic localisation using smartphones is even possible in 3D space, as shown by Qiu et al. in [8], using two microphones per phone in an unobstructed low-noise environment.

However, most of these approaches are hardly applicable to the case where both phones are carried in the persons' pockets, as the noise caused by the phone rubbing against the fabric during movement is dwarfing most signals. Instead of trying to measure the distance, we focus on detecting whether two persons are standing proximate to each other or encountering each other through walking by.

III. The Algorithm

Experimenting with the speakers of an iPhone 4 and aiming for minimal annoyance of users, we discovered that the frequency range of 18–23 kHz – being inaudible to most people – is applicable for sound emission as well as detection. The phone's speaker is able to produce sound in that range and the microphone's frequency response is fairly linear and big enough to be measured; our findings are consistent with [6].

Consider two devices D_A and D_B being in the pockets of two people walking relatively distant, but approaching each other. As soon as D_A 'sees' D_B via Bluetooth, it connects to D_B, they negotiate a carrier frequency f and assign the roles of *sender* and *receiver*. Using these two roles is a concession to the iOS-platform, as it turned out that simultaneous playback and recording is artificially limited in playback volume by the OS and the delays introduced by dynamically switching between these modes were too high to be considered a viable option.

The sender starts emitting a sine wave oscillating at the negotiated frequency f. Sending does not take place continuously rather than being amplitude modulated. Amplitudes are determined by a certain pattern containing k binary entries, which represent intervals of length l_s, the length of one pattern is $l_p = k \cdot l_s$. This approach is closely related to On-off keying (OOK), although in contrast to data transmission, we aim at detectability of the signal at all.

The receiver records the ambient sound and performs a Fast Fourier Transform (FFT) for every interval of length l_s of the input signal. Figure 1 shows the ideal pattern $[1, 1, 0, 1, 0, 0]$, which is being sent on a carrier frequency of 18 kHz, overlaid with the magnitudes of the corresponding FFT result for f of the signal received under ideal circumstances. We chose a sampling rate of 44.1 kHz and intervals of 512 samples, which makes the interval length $l_s = \frac{44,100}{512} \approx 0,01161s$ and with $k = 6$ values per pattern, the pattern length is $l_p \approx 0,07$s. We decided to use a rather short pattern, as this allows for shorter window length and hence faster detection, while increasing the possibility to 'slip through' in low-noise phases. Furthermore, variance in signal amplitude per window is lower using shorter windows in case the persons move. The FFT has been chosen over a finite impulse response filter (FIR) because it allows us to analyze several frequencies (i.e. another sender) at once at little additional cost. Furthermore, we leverage the fact that iOS offers high-performance vector-based functions for performing FFTs.

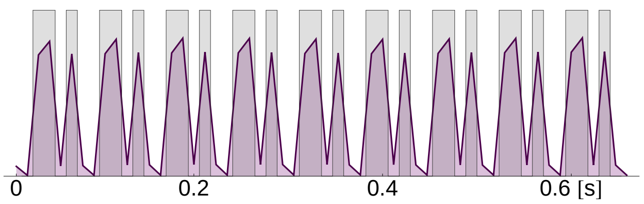


Fig. 1. An ideal sequence of patterns, overlaid with the recorded signal when speaker and microphone face each other in 10 cm distance.

As the FFT yields complex results and we are only interested in the signal's amplitude, we compute their magnitude $m = |z_f|$ for each FFT bin $z_f = (x, y) \in \mathbb{C}$ where f is a frequency of our interest. The same is being done for an adjacent but unused frequency, which we take as a sample of the overall noise level in that frequency range.

Noise, in this case, is mainly created by the phone rubbing against the pocket's fabric and is both, extremely loud and quite evenly distributed across adjacent frequencies. Figure 2 shows samples of frequency spectra as recorded by the receiver with both phones in different positions. As one can easily see, the signal in line of sight (see Fig. 2(a)) is already several orders of magnitude louder than the sample with both phones in the pocket and standing still (see Fig. 2(b)), which is still excellent compared to Fig. 2(c), in which the signal is literally drowned in noise created by rubbing against the pocket. This kind of noise poses the most severe challenges to our detection approach.

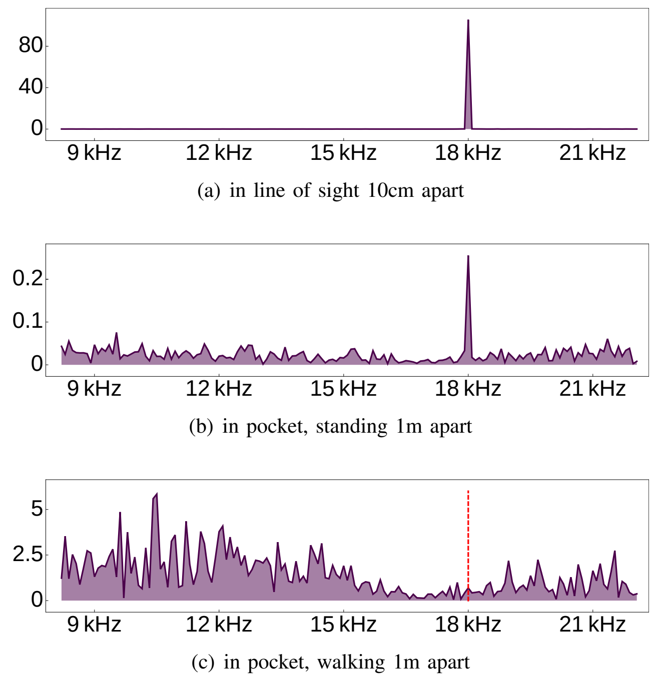


Fig. 2. Samples of frequency spectra as recorded by the receiver, while both phones are positioned as stated in (a) – (c)

Each k magnitudes of the carrier frequency as well as the frequency used as a noise sample are considered one window and are cross-correlated with the window earlier in time, yielding the correlation coefficients c_s for the signal and c_n for noise, respectively. The beauty of this approach lies in auto-correlating windows of length l_p, which are known to contain exactly one pattern, albeit offset by an unknown amount of time. Consequently, synching the receiver with the sender can be omitted and clock drift is mitigated, as it is quite small between two windows.

We create a new correlation measure $c_d = c_s - c_n$, which approximates the amount by which our signal outgrows the noise. Subsequently, the signal auto-correlation c_s as well as

c_d are averaged using sliding windows of different sizes to account for different characteristics:

- a_d is the sliding average of c_d over a larger window of $k_{a_d} = 16$ values in order to statistically collect traces of the signal in noise, while bursts of noise yielding small or negative values should even out over time
- a_s is the sliding average of c_s over a smaller window of $k_{a_s} = 8$ values, which – if high – represents a fairly low-noise situation in the short term

These two averages are subject to a simple thresholding in order to cover two scenarios, for which $t_H > t_L$:

1) A high-noise scenario, typically arising while walking. A detection candidate satisfies the following condition:
$$d_1 = (a_d > t_H) \vee (a_s > t_H),$$
which enforces candidates to have a relatively high auto-correlation despite noise in the long term *or* a relatively high autocorrelation in the short term, typically in between two steps.
2) A low-noise scenario, typically arising while not moving. A detection candidate satisfies the following condition:
$$d_2 = (a_d > t_L) \wedge (a_s > t_L),$$
which enforces candidates to have a moderate auto-correlation despite noise in the long-term *and* a moderate autocorrelation in short-term, which is the case when standing still or sitting, even when the signal is weak and mixed with moderate ambient noise.

The signal is considered partially detected if $d_p = d_1 \vee d_2$ is true. However, in the top-most abstraction layer, each 14 (because $l_p \cdot 14 \approx 1\text{s}$) values of d_p are counted and required to exceed yet another threshold, tailored for the two investigated situations: standing and walking. Constants are chosen carefully to almost entirely rule out false positives as well as successful detections too far away to be considered proximate.

IV. Evaluation

Overall, we have collected 137 minutes of audio data (thereof 64 minutes without emitting a signal for estimating algorithm's precision) for the standing and walking scenarios as listed in Table I and Table II. Phones were carried in jeans pockets, with no extra cases. The parameters of the program were chosen with an ambitious goal of delivering reliable results, with no time lag, high resolution (up to 2.5m), live on the device. In particular, we considered False Positives (which according to our definition included also proximity detection of people standing more that 3m apart) much more harmful than False Negatives (i.e. not detecting proximity) and thus, at the price of a slightly lower recall rate, we set the bar for the precision of the algorithm very high.

With auto-correlation thresholds $t_L = 0.355$, $t_H = 0.42$, and 9 out of 14 d_p-values for a positive detection (translating to a time resolution of approximately one second), the True Negative rate amounted to 99.95%, meaning a nearly perfect precision of the algorithm.

TABLE I
Standing scenario (indoor & outdoor). Recall rates for 1s and 10s window (majority-decision). In a 10s variant, distances up to 2m are well recognized, proximity of 4m and greater is considered being too far away, as desired.

Activity	**Dist. [m]**	**Recall (1s)**	**Recall (10s)**
Office corridor	0.6	0.95	1
Shopping mall escalator	1	0.71	1
Supermarket queue	1	0.53	0.60
Office corridor	1.2	0.94	1
Busy street crossing	1.2	0.80	1
Busy street crossing	2	0.22	0.25
Neighbouring rooms	2	0	0
Office corridor	2.5	0.38	0.31
Shop shelf in between	3	0.06	0.11
Busy street crossing	3	0.01	0
Office corridor	4	0.14	0
Busy street crossing	4	0	0
Busy street crossing	5	0	0
Office corridor	6	0.09	0
Supermarket	6	0	0
Office corridor	10	0	0

First, we investigated the social interaction scenario, i.e. a situation, when two people are standing next to each other. For this static case we used a majority-decision window of 10 seconds (which is not too long, considering nobody is moving). The results are presented in Table I and Fig. 3. Figure 3 shows the proximity detection rate as a function of distance between the two persons and the majority threshold cut-off for the 10-second recognition window.

In overall, the algorithm almost perfectly recognized proximity up to 2m (both indoor & outdoor), while situations when people were more than 3m apart were correctly classified as non-proximate.

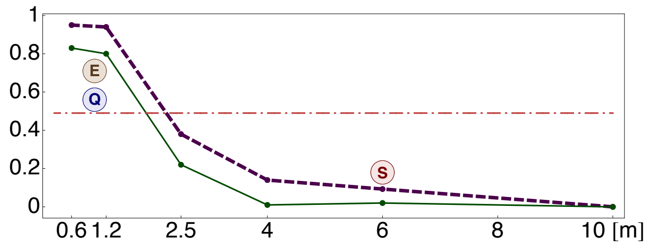


Fig. 3. Overall recall rates as a function of distance with both persons standing: indoors (dashed), near a busy street crossing (solid), Ⓔ on an escalator, Ⓠ in a supermarket queue, Ⓢ remotely at different supermarket shelves. Hortizontal line corresponds to the majority-decision window of 10 seconds (see Table I).

Next, we investigated the walking scenario. In this case, when the receiver starts walking, the signal-to-noise ratio (SNR) drops from 9.5 dB to 2 dB, thus making it extremely hard to cover both, standing and walking, with a fixed set of parameters. This can also be seen in Fig. 4, which depicts the average detection rate of 11 recordings increasing every second as the sender approaches the standing receiver, bearing a peak of nearly 100% when closest and more rapidly falling than rising due to acoustic shielding of the leaving person's body. With the receiver walking, on the other hand, the detection rate both, increases at a slower rate and the duration of subsequent detections is significantly lower.

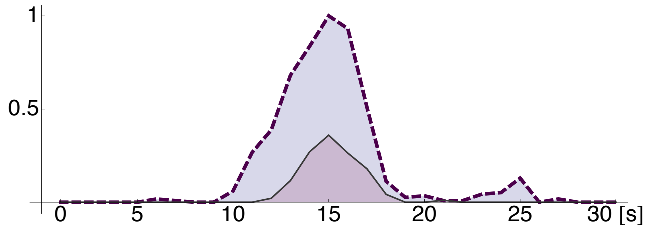


Fig. 4. Average detection rates with one person standing: (a) sender walking past receiver (dashed) and (b) receiver walking past sender.

Results for the walking scenario are shown in Table II. As anticipated, there is a big discrepancy in the recall rates between the 'standing receiver' and the 'walking receiver' scenario. One of the reasons of the 0.33 recall rate for the walking receiver is a very high bar on the precision rate of the algorithm.

Although the walking scenario poses a major challenge due to the low signal-to-noise ratio, the is still room for improvement. We plan swapping the sender and the receiver role every few seconds (and thus be able to detect proximity if at least one person is static).

TABLE II
WALKING IN A SPORTS SHOP: SENDER WALKING TOWARDS AND AWAY OF THE RECEIVER (11 RECORDINGS), RECEIVER WALKING TOWARDS AND AWAY OF THE SENDER (9 RECORDINGS).

Activity	**Precision**	**Recall**
Receiver Standing	1	1
Receiver Walking	1	0.33

V. CONCLUSION

In this paper we have presented an algorithm which uses inaudible sound patterns to accurately detect whether two mobile phones are within few meters from each other. The algorithm is implemented as a standalone iPhone application and is fast enough to be run live on the device. Tests in a variety of environments confirm the robustness of the method.

In authors' view, an extremely promising application of the results lies in the collaborative localization domain [2]. Indoor positioning algorithms running on the mobile phones (e.g. algorithms using motion sensors for inertial navigation) could leverage proximity information to constantly calibrate their location estimates during social encounters.

Future work needs to look at improving recognition rates when both users are walking and, in order to make the deployment of the algorithm on a bigger scale feasible, address the scalability issue, when dozens of devices are within Bluetooth range. As for the former, using wider frequency ranges and more complex codings may be another option to improve performance. With some phones having two microphones for noise canceling, differential approaches may be used to get rid of noise generated by the phone rubbing on the fabric when the user is moving.

ACKNOWLEDGEMENT

This work is supported by the FP7 ICT Future Enabling Technologies programme of the European Commission under the grant agreement No. 231288 (SOCIONICAL).

REFERENCES

[1] J. A. Paradiso, J. Gips, M. Laibowitz, S. Sadi, D. Merrill, R. Aylward, P. Maes, and A. Pentland, "Identifying and facilitating social interaction with a wearable wireless sensor network," *Personal Ubiquitous Comput.*, vol. 14, no. 2, pp. 137–152, Feb. 2010. [Online]. Available: http://dx.doi.org/10.1007/s00779-009-0239-2

[2] K. Kloch, P. Lukowicz, and C. Fischer, "Collaborative PDR Localisation with Mobile Phones," in *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers*, ser. ISWC '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 37–40. [Online]. Available: http://dx.doi.org/10.1109/ISWC.2011.16

[3] M. Wirz, D. Roggen, and G. Tröster, "A wearable, ambient sound-based approach for infrastructureless fuzzy proximity estimation," in *Proceedings of the 14th IEEE International Symposium on Wearable Computers (ISWC 2010).* IEEE Computer Society, Oct. 2010.

[4] J. Scott and B. Dragovic, "Audio location: accurate low-cost location sensing," in *Proceedings of the Third international conference on Pervasive Computing*, ser. PERVASIVE'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 1–18. [Online]. Available: http://dx.doi.org/10.1007/11428572_1

[5] L. Girod, M. Lukac, V. Trifa, and D. Estrin, "The design and implementation of a self-calibrating distributed acoustic sensing platform," in *Proceedings of the 4th international conference on Embedded networked sensor systems*, ser. SenSys '06. New York, NY, USA: ACM, 2006, pp. 71–84. [Online]. Available: http://doi.acm.org/10.1145/1182807.1182815

[6] W. A. Arentz and U. Bandara, "Near ultrasonic directional data transfer for modern smartphones," in *Proceedings of the 13th international conference on Ubiquitous computing*, ser. UbiComp '11. New York, NY, USA: ACM, 2011, pp. 481–482. [Online]. Available: http://doi.acm.org/10.1145/2030112.2030181

[7] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *Proceedings of the 5th international conference on Embedded networked sensor systems*, ser. SenSys '07. New York, NY, USA: ACM, 2007, pp. 1–14. [Online]. Available: http://doi.acm.org/10.1145/1322263.1322265

[8] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '11. New York, NY, USA: ACM, 2011, pp. 190–203. [Online]. Available: http://doi.acm.org/10.1145/2070942.2070962