

Micrófonos de contacto: una alternativa para sensado táctil en robots sociales

Juan José Gamboa
University Carlos III of Madrid
jgamboa@ing.uc3m.es
Sara Marqués Villaroya
University Carlos III of Madrid
smarques@ing.uc3m.es

Fernando Alonso-Martín
University Carlos III of Madrid
famartin@ing.uc3m.es
Marcos Maroto
University Carlos III of Madrid
marmarot@pa.uc3m.es

José Carlos Castillo
University Carlos III of Madrid
jocastil@ing.uc3m.es
Miguel Ángel Salichs
University Carlos III of Madrid
salichs@ing.uc3m.es

Resumen

En la interacción humano-máquina, especialmente en la robótica social, el contacto directo entre el humano y el robot, y todo lo que permite al robot reaccionar y obtener información de este contacto, tiene una gran relevancia. Entre la información que el robot obtiene se encuentra aquella relacionada con la parte del cuerpo del robot que se está tocando y qué gesto se pretende transmitir con dicho toque. A través de este artículo, se pretenden evaluar, en primer lugar, las actuales tecnologías que envuelven la interacción táctil aplicada a la robótica. Por otro lado, se busca aportar un enfoque nuevo a la interacción táctil en la robótica social, presentando aplicaciones y sistemas que emplean el sonido como fuente de información principal. Este trabajo propone un sistema de detección de gestos táctiles en un robot social empleando tecnologías de adquisición sonora.

Keywords— percepción acústica, interacción táctil, interacción humano-máquina, robótica social

1. Introducción

A través del tacto los seres humanos somos capaces de transmitir una gran cantidad de información. Desde un cálido apretón de manos a un abrazo, hemos desarrollado lenguajes complejos y expresiones emocionales a través del contacto físico [1]. Se pretende dotar, del mismo modo, a un robot social de estas habilidades, ya que multiplican el número de posibilidades por las cuales se podría interactuar con él [2], añadiendo una fuente de información que, como se ha indicado antes, posee una gran cantidad de matices. El problema que surge en la actualidad es que los sistemas que se encuentran disponibles requieren cubrir prácticamente la totalidad de la carcasa del robot con sensores [3].

Un robot social necesitaría de un modo que permita percibir los estados emocionales del usuario, o simplemente poder comprender de una mejor manera la intención comunicativa del usuario con respecto al robot. El tacto es un tipo de interacción

que permite la transmisión de esta información. El problema a la hora de implementar este tipo de sistemas es elegir el tipo de tecnología más adecuado. En este caso se propone el sonido como vehículo para esta transmisión de datos. La acción de tocar una superficie sólida genera un contacto que puede ser percibido por distintos tipos de tecnologías, pero además un sonido se genera. Éste puede ser detectado, procesado y reconocido empleando técnicas de análisis de audio y micrófonos, obteniendo además información con una gran cantidad de matices, ya que es una señal que posee características en varios dominios, como son el del tiempo el de la frecuencia y el del tiempo-frecuencia. Esta perturbación se esparce fácilmente por los sólidos, pero no tanto en el aire. Esto implica que colocando el sensor adecuado sobre una superficie se podrá obtener esta señal con una gran precisión, buscando además ignorar los ruidos que se propaguen por el aire, como la voz o cualquier sonido ambiental. Estos dispositivos se conocen como micrófonos de contacto.

A lo largo de este trabajo se exploran las principales tecnologías que en la actualidad, en el campo de la interacción humano-máquina, se han desarrollado con el objetivo de identificar distintos tipos de contacto a través de varias clases de sensores, aplicando además aprendizaje automático. También se presentan diversas aplicaciones que, en otros campos, emplean tecnología relacionada con el sonido para reconocer gestos táctiles. Este trabajo propone, a través de la revisión de sistemas que implementan interacción táctil en robótica y de interfaces táctiles basadas en el análisis de una señal de audio, un sistema que emplee micrófonos de contacto para detectar gestos táctiles en un robot. Este sistema se complementará con diversos algoritmos de aprendizaje automático, con el objetivo de caracterizar e identificar el gesto detectado.

2. Tecnologías actuales

En la siguiente sección se presentan las tecnologías actuales relacionadas con la interacción táctil a través de dos perspectivas. En primer lugar se introducen distintas tecnologías relacionadas con la interacción humano-máquina táctil, tecnologías

que involucran el uso de sensores de fuerza, de visión o incluso de temperatura, además de aprendizaje automático, para encontrar patrones en la información que estos sensores perciben, de forma que se pueda identificar el contacto que se está realizando.

La otra perspectiva que se plantea es la de diferentes interfaces táctiles que involucran sonido. En este caso no se enfocan en la interacción humano-máquina en sí, sino que son aplicaciones con propósitos más diversos, como la comunicación con un servicio contestador automático o el control de un dispositivo de audio.

2.1. Interacción humano-máquina táctil

En el trabajo de Argall [3] encontramos un estudio completo acerca de las habilidades táctiles aplicadas a la robótica social que compara distintos robots con materiales diferentes en sus carcasas. En el caso de robots con superficies duras, los sensores que se suelen emplear son los siguientes: fuerza/torque, Resistencias Sensibles a Fuerza (FSR), acelerómetros y sensores de deformación. Por otro lado, en los robots de “piel blanda”, se suelen incorporar: potenciómetros (información kinestética), FSR’s, sensores de capacitancia, sensores de temperatura (termistores), sensores de campo eléctrico, y fotorelectores. Argall no menciona el uso de micrófonos como base para habilidades de tacto. En general, los sensores que se suelen usar en HRI tienen las siguientes deficiencias: requieren prácticamente contacto directo y su alcance para detectar contacto está limitado a unos pocos centímetros, por tanto es necesario equipar al robot con una gran cantidad de sensores para cubrir prácticamente por completo la carcasa del mismo (en la mayoría de los casos son más de 40 sensores por robot). Por ejemplo, el robot “abrazable” presentado por Stiehl integra más de 1000 sensores de fuerza, 400 sensores de temperatura, y 9 sensores de campo eléctrico [4].

En posteriores trabajos, Silvera lleva a cabo múltiples experimentos [5, 6] empleando un sistema compuesto por un brazo artificial cubierto por una piel sensitiva basada en tecnología de impedancia eléctrica tomográfica que es capaz de diferenciar entre múltiples tipos de tacto. Para ello, los autores incluyeron un algoritmo de aprendizaje automático denominado *LogicBoost* [7] con el que se logra reconocer seis tipos de contacto. Con esta configuración, Silvera afirma que obtiene el 80% de precisión, similar a la que poseen los seres humanos.

Siguiendo la misma línea, Cooney presenta un trabajo que intenta diferenciar distintos contactos afectivos sobre el robot humanoide Kirin [8]. Pa-

ra clasificar los toques, construye un clasificador *SVM* (*Support Vector Machine* en inglés) y un clasificador *k-NN* (*k* vecinos más cercanos). El trabajo descrito está basado en técnicas de visión con sensores Kinect, y 14 sensores KINOTEX¹ modificados [9] y distribuidos sobre la superficie del cuerpo del robot. Mezclando los resultados de los clasificadores, afirma obtener un 90.5% de precisión.

2.2. Interfaces táctiles basadas en el análisis de sonido

Hasta este punto, se ha analizado el estado del arte en el campo de la interacción humano-máquina asociada a la robótica. Esta sección está dedicada al análisis de las interfaces que incorporan tecnología de audio para detectar y/o clasificar toques dentro de diversas áreas tecnológicas no relacionadas con la interacción humano-máquina (HRI, por sus siglas en inglés), debido a que hasta ahora no hay abundantes aplicaciones en el campo de la interacción humano-máquina.

Paradiso [10] describe un sistema para localizar de un modo brusco la posición de golpes y toques sobre una superficie de cristal de un tamaño de un metro cuadrado. Paradiso emplea cuatro receptores piezoeléctricos (conocidos también como micrófonos de contacto) colocados en las cuatro esquinas del cristal. Específicamente usa el modelo *Panasonic EFV-RT series*. Paradiso afirma en el informe obtener grandes resultados, con una precisión de entre 2 y 4 centímetros.

Harrison [11] perfiló un sistema que permite, a través del análisis de audio, reconocer toques sobre distintos materiales, especialmente escritorios y muros. Afirma que este sistema puede diferenciar entre seis tipos de toques con una precisión alrededor del 90%. Para ello, emplea un estetoscopio modificado que es capaz de amplificar la señal recibida. En este trabajo, el autor no intenta localizar la fuente del sonido, se centra principalmente en reconocer el tipo de toque realizado.

Murray-Smith [12] introduce *Stane*, un dispositivo electrónico controlado por interacción táctil basada en micrófonos. Específicamente, *Stane* puede reconocer esos gestos: *stroke* (caricia), *scratch* (arañazo), *rub* (frote) y *tap* (toque con un dedo). Para ello, usa una combinación de sensores: ca-

¹consisten en un diodo emisor de luz, un fotodetector, espuma de uretano, y cables de fibra óptica. Los cables se usan para irradiar a la espuma con un rayo de luz fino, y para recoger la luz esparcida por la espuma. El principio de funcionamiento de los sensores KINOTEX es medir a través del fotodetector cambios en la intensidad de la luz al final del cable de fibra óptica

pacitivos, inerciales, y micrófonos de contacto. La aplicación real de *Stane* es controlar un reproductor musical. Murray-Smith argumenta que el uso de micrófonos de contacto o piezo-micrófonos es perfectamente adecuado para recibir las vibraciones generadas por ese tipo de contactos sobre superficies sólidas. Para la tarea de clasificación, emplea una red neuronal entrenada con 26880 ejemplos. Sus resultados experimentales muestran una tasa de acierto en torno al 75% para los cinco tipos de contacto indicados anteriormente.

Lopes [13] muestra un sistema que intenta extender los sistemas multi-toque tradicionales. Mezcla dos tecnologías distintas, sensores capacitivos para detectar la posición del toque, y percepción acústica para reconocer diferentes gestos táctiles como: dedos, nudillos, uñas de los dedos y puños. Aunque Lopes muestra una interacción bastante precisa, la publicación no indica específicamente resultados sobre la tasa de acierto.

2.3. Análisis de las tecnologías y trabajos estudiados

Como se mostraba en el análisis de Argall, la introducción de sensores para realizar tareas de interacción táctil implica costosas redes sensoriales (más de mil sensores en el sistema que presenta). Silvera corrobora esto pero planteando una serie de gestos táctiles de gran interés. Por último Cooney introduce el uso de algoritmos como máquinas de soporte vectorial (SVM, por sus siglas en inglés) y los k vecinos más cercanos (k -NN).

En cuanto a la aplicación de tecnología acústica, en la mayor parte de casos es empleada para localización directa del origen, procesando la señal obtenida por los micrófonos. En el caso de *Stane*, Murray-Smith aplica clasificación (redes neuronales) para detectar gestos. Lopes presenta un trabajo similar, pero sin dar detalles de la fase de clasificación.

A partir de las ideas que se muestran en estos dos enfoques, en este trabajo se plantea un sistema que, aplicado a la interacción humano-máquina, combine conceptos de las dos tecnologías que se acaban de analizar. En cuanto a los sensores que se deciden emplear, se han seleccionado los micrófonos de contacto, siendo uno de los objetivos que el sistema evite emplear una gran cantidad de sensores. Gracias a las ideas de Murray-Smith, con conjunto con las de Silvera, se muestra un atisbo de lo que sería dirigir la robótica y los micrófonos de contacto hacia la detección de gestos táctiles. Por último, la mayor parte de los trabajos insertan un bloque de clasificación que permita sacar conclusiones de la información obtenida, destacando

clasificadores como *SVM*, *k-NN* y redes neuronales.

3. Sensado táctil en un robot social

La implementación de un sistema que use micrófonos de contacto para detección de toques consiste en 4 fases que se van a desglosar a lo largo de este capítulo: Instalación de los micrófonos de contacto, detección de la actividad táctil, extracción de las características de la señal obtenida durante la actividad y por último, la clasificación.

3.1. Instalación de los micrófonos

La selección de micrófonos de contacto está basada en la calidad de la señal captada y la sensibilidad que poseen en superficies lisas y rígidas. Otra de las características es que aísla de manera natural ruidos que se propaguen por el aire, lo que hace que los ruidos externos más corrientes, como la voz humana no sean recogidos por el sensor. La desventaja principal que se observa en un primer momento es que, precisamente esa fidelidad en la extracción de las vibraciones que se propagan por los sólidos, causa que los ruidos internos dentro del robot provocados por los servomotores puedan causar falsos positivos. En cualquier caso eso sería un problema abordable tanto desde el punto de vista del hardware, introduciendo aislamiento físico, como desde el software, a través de filtros en el extractor de características o en el clasificador.

De entre los posibles tipos de micrófonos nos decantamos por uno de precio asequible (35 euros) y de una calidad bastante contrastada por profesionales del mundo de la música. El modelo elegido ha sido el *Oyster Schaller 723* mostrado en la figura 1.



Figura 1: Micrófono de contacto Oyster Schaller 723 S/S

El robot social que ha servido como plataforma de investigación, y por lo tanto, como base para realizar los experimentos realizados, es el robot Maggie, del Robotics Lab de la Universidad Carlos III de Madrid[14].

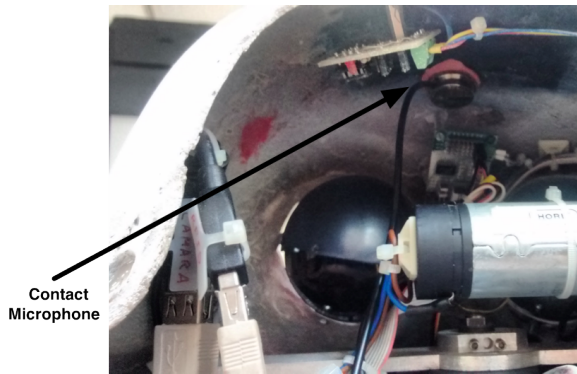


Figura 2: Micrófono instalado en el interior de la cabeza de Maggie

En la primera versión se decide instalar un total de tres micrófonos en las zonas que se consideran que se va a establecer con más probabilidad un contacto. Estas partes son la cabeza, y el pecho (izquierda y derecha). Para fijar los micrófonos en el interior de la carcasa, que es rugosa y cóncava se ha empleado arcilla para modelar y adhesivo de dos componentes. Gracias a las propiedades moldeables pero rígidas al secarse de ambos, se puede fijar la totalidad de la superficie sensible del micrófono al interior del robot, como se muestra en la Fig. 2. Los micrófonos serán conectados en Maggie de la siguiente manera. La salida Minijack es conectada a unas tarjetas de sonido USB SoundBlaster² de alta fidelidad.

3.2. Detección de actividad táctil

A nivel de software se ha empleado ChucK [15], un lenguaje de programación diseñado para proveer a los usuarios de una plataforma para síntesis y análisis precisa de audio.

Para la detección como la extracción de características de la señal sonora se ha empleado una versión modificada para esta tarea de GEVA (*Gender and Emotion Voice Analysis*) [16], un programa creado para detectar tanto el género como la emoción que transmite un interlocutor al hablar. GEVA es un sistema de análisis de sonido diseñado y desarrollado en ChucK que combina varios elementos pertenecientes a tres dominios diferentes: tiempo, frecuencia y tiempo-frecuencia. La principal funcionalidad del programa es que calcula las características pertenecientes a cada uno de los dominios en tiempo real, analizando directamente la señal sonora capturada.

GEVA presenta una serie de umbrales ajustables referentes a la señal que permiten diferenciar cuando parte del sonido recibido, ya que los micrófo-

nos se encuentran escuchando continuamente, no es ruido ambiental o un tipo específico de sonido que se quiera aislar, y marca el principio de un gesto. Uno de estos umbrales, y el que se emplea principalmente es el *Signal-to-Noise Ratio*, que se explicará más adelante.

3.3. Extracción de características

Una vez que el sistema considera que el gesto ha comenzado, GEVA mide distintas características, englobadas en tres dominios diferentes, pero relacionadas entre sí mediante transformadas.

- Las características que se encuentran en el *dominio del tiempo* se obtienen directamente del muestreo de la señal analógica que proviene del micrófono.
- En el caso de las características que están en el *dominio de la frecuencia* se aplica la transformada rápida de Fourier (FFT) sobre la señal en el dominio del tiempo.
- Por último, para obtener las características pertenecientes al *dominio del tiempo-frecuencia* se aplica la transformada discreta de Haar (*Discrete Haar Wavelet Transform*).

Las características que se obtienen de la señal son los siguientes.

1. **Pitch:** Es una propiedad de los sonidos que permite su ordenación en una escala relativa a la frecuencia [17]. Cuantitativamente, las estimaciones del pitch son medidas del periodo de oscilación y se pueden realizar en los tres dominios mencionados anteriormente. En cada dominio se encuentran implementados distintos algoritmos de detección de pitch (PDA o pitch detection algorithms):
 - a) El método para detectar el primer mínimo en la autocorrelación de la señal de entrada en el dominio del tiempo. Esto implica que el pitch se corresponde al mínimo en la relación de cruce entre la señal sonora consigo misma [18, 19].
 - b) El segundo método aplica la transformada rápida de Fourier. Este algoritmo considera la distancia entre los picos en el espectro de la señal para calcular el pitch.
 - c) El último método para obtener el pitch se desarrolla en el dominio del tiempo-frecuencia, y usa la transformada discreta de Haar para ondas. Este cálculo se basa en el trabajo de Larson [20].

²Diseñadas por Creative Labs. Modelo SB1300.

2. **Flujo**: El flujo indica si hay grandes variaciones en el dominio de la frecuencia. Cuando toma valores cercanos al cero implica que los valores de amplitud obtenidos en todos los rangos de frecuencia son muy similares.
3. **Rolloff-95**: Se corresponde con el valor de frecuencia que contiene el 95 % de la energía de la señal.
4. **Centroide**: Representa la mediana del espectro de la señal. Esto es la frecuencia a la que la perturbación se aproxima en mayor medida. Se usa frecuentemente para calcular el tono de un sonido o la voz (timbre).
5. **Zero-crossing rate (ZCR)**: Indica el número de veces que la señal cruza el cero (en el eje x). Normalmente, el sonido de fondo cruza con mucha más frecuencia el eje x que los sonidos verbales.
6. **Signal-to-noise ratio (SNR)**: Permite relacionar el sonido que genera el gesto con respecto al volumen del ruido de fondo.
7. **Root Mean Square (RMS)**: relativo a la amplitud del volumen de la señal.
8. **Duración**: hace referencia a duración del sonido.

Cuando finaliza el contacto, el programa emite los valores máximos, mínimos, valores medios y la diferencia entre máximo y mínimo de las características descritas.

3.4. Clasificación automática

Entre las librerías disponibles para realizar tareas relacionadas con el aprendizaje automático, se ha seleccionado *Weka* por su versatilidad. *Weka* soporta varias tareas estándar de minería de datos, especialmente, preprocesamiento de datos, *clustering*, clasificación, regresión, visualización, y selección. Todas las técnicas de *Weka* se fundamentan en la asunción de que los datos están disponibles en un fichero plano (flat file) o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos).

Weka integra una amplia familia de clasificadores³ pero entre ellos destacan los que han sido empleados en trabajos como el de Cooney o el de Murray-Smith y los árboles de decisión:

- **Random Forest**. Familia de algoritmos predictivos que emplean la técnica de *bagging* o empaquetado para combinar diferentes árboles de decisión, donde cada árbol es construido con observaciones y variables aleatorias [21]. El algoritmo funciona como se explica a continuación.

En primer lugar, selecciona individuos al azar (usando muestreo con reemplazo) para crear diferentes set de datos. Posteriormente crea un árbol de decisión con cada set de datos, obteniendo diferentes árboles, ya que cada set contiene diferentes individuos y diferentes variables. Al crear los árboles se eligen variables al azar en cada nodo del árbol, dejando crecer el árbol en profundidad (sin podar). Y por último, predice los nuevos datos usando el “voto mayoritario”, donde clasificará como “positivo” si la mayoría de los árboles predicen la observación como positiva.

- **k-NN**. Este es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad o directamente la probabilidad a posteriori de que un elemento x pertenezca a la clase C_j a partir de la información proporcionada por el conjunto de modelos [22]. En el proceso de aprendizaje no se hace ninguna suposición acerca de la distribución de las variables a predecir. En el reconocimiento de patrones, el algoritmo es usado como método de clasificación de objetos (elementos) basado en un entrenamiento mediante ejemplos cercanos en el espacio de los elementos. *k-NN* es un tipo de *Lazy Learning*, donde la función se aproxima solo localmente y todo el conjunto de cálculos es aplaza hasta la clasificación.

- **SVM**. Weka internamente implementa la optimización secuencial mínima o SMO, un algoritmo para resolver el problema de programación cuadrática que surge durante el entrenamiento de *support vector machines* [23]. SMO es un algoritmo iterativo para resolver este problema de optimización. Divide el problema principal en una serie de subproblemas lo más pequeño posible, que son resueltos analíticamente.

- **Perceptrón Multicapa**. Un clasificador basado en una red neuronal que utiliza *backpropagation* para clasificar instancias. La red se puede construir a mano, crearse a través de un algoritmo o de ambas formas. Ésta también puede ser monitorizada y modificada durante el tiempo de entrenamiento. Los nodos de esta red son todos sigmoides (excepto cuando la clase es numérica, en cuyo caso los nodos

³<http://weka.sourceforge.net/doc/stable/weka/classifiers/package-frame.html>

	Área de contacto	Intensidad	Duración	Intención	Ejemplo
Caricia	media-grande	baja	media-larga	empatía, compasión	
Cosquillas	media	media	media-larga	diversión, alegría	
Toque	pequeña	baja	corta	llamada, aviso	
Cachete	pequeña-media	alta	corta	disciplina, castigo	

Tabla 1: Caracterización de los gestos empleados para las pruebas. La última columna muestra un ejemplo de cómo realizar dicho gesto.

de salida se convierten en unidades lineales no desalineadas).

Esta parte comprende la creación de un conjunto de instancias preclasificadas que se irán creando en experimentos con usuarios. Cuando el conjunto de instancias sea lo suficiente grande se integrará en el sistema para poder entrenar a los clasificadores. De este modo, los clasificadores podrán ir clasificando de modo *online* las nuevas instancias desconocidas que vayan recibiendo.

3.5. Resultados preliminares

Para evaluar el funcionamiento del sistema propuesto se realizó una serie de pruebas para medir la tasa de acierto. En primer lugar se han definido una serie de gestos mostrados en la tabla 1, inspirados por los experimentos de Silvera, previamente explicados. Tras esto, se ha procedido a obtener muestras de entrenamiento. El conjunto de muestras posee las siguientes características. Se han recogido un total de 747 interacciones con el robot, realizadas por un total de 10 usuarios. De los cuales 215 son *caricias*, 48 son *cosquillas*, 260 son *toques*, y 224 son *cachetes*.

Las interacciones entre el robot y los usuarios se han realizado de la siguiente forma. El supervisor pide al usuario que toque al robot en la cabeza usando los cuatro posibles gestos a reconocer. Cada uno de estos tipos de toques los puede repetir un número indeterminado de veces. Cada uno de los sonidos producidos en cada interacción queda registrado por el sistema y es etiquetado con el nombre del gesto realizado por el usuario.

Clasificador	<i>F-score</i>
Random Forest	1
Perceptrón Multicapa	0.93
SVM	0.80
k-NN	0.78

Tabla 2: F-Score de los clasificadores empleando validación cruzada.

Clasificador	<i>F-score</i>
Random Forest	0.79
Perceptrón Multicapa	0.75
SVM	0.72
k-NN	0.66

Tabla 3: F-Score de los clasificadores empleando un conjunto de prueba.

En la fase de entrenamiento, los datos se almacenan en un fichero que posteriormente serán usados tanto para entrenamiento como para validación. En la fase de validación se han empleado dos técnicas: Validación cruzada, que se ha hecho con subgrupos de los datos de entrenamiento de 10 instancias cada uno. Adicionalmente, se recogieron nuevos datos de 15 usuarios diferentes a los 10 que realizaron las muestras previas. En las tablas 2 y 3 se muestra el *F-Score* resultante para cada algoritmo expuesto previamente. Se trata de un indicador del grado de acierto de un determinado clasificador en base a 1, que representa el acierto en todas las muestras.

4. Conclusiones

El sonido es un tipo de señal de la que se pueden obtener una gran cantidad de características, lo que provee a la información transmitida de muchos matices. Esto facilita que la implementación de algoritmos de clasificación aporte muy buenos resultados. Los micrófonos de contacto son un tipo de receptor de audio que demuestra baja distorsión y ruido, percibiendo las vibraciones con una gran precisión. Estos micrófonos son capaces de aislar los ruidos ambientales que se propagan por el aire, que suelen ser los más corrientes (voces, etc). Debido a las propiedades de transmisión del sonido de las superficies sólidas, los ruidos internos se propagan con facilidad y puedan ser recibidos, creando falsos positivos. Como se muestra en el montaje, los micrófonos de contacto tienen como limitación física las superficies rugosas e irregulares y las blandas.

En el trabajo se ha aplicado el aprendizaje como una herramienta capaz de aprovechar la gran riqueza que se obtiene de una señal sonora. Pero dentro de este campo hay protocolos y herramientas que permiten mejorar y profundizar en esta fase del sistema (p. ej. meta-clasificadores), lo que permitiría un margen de mejora muy amplio. Además, con una gran cantidad de información de entrenamiento se pueden aplicar técnicas más complejas pero con mayor tasa de acierto.

Actualmente la tecnología de adquisición de sonido engloba dispositivos que son capaces de obtener una gran cantidad de información de una señal que, como en este caso, se propaga por los sólidos (superficie de un robot). Esto, junto al gran impulso que en la actualidad está viviendo el aprendizaje automático y la creciente capacidad de procesamiento de las computadoras, provoca que la tecnología que se presenta en este trabajo constituya una importante vía para distintos sistemas sensoriales aplicados a la interacción humano-máquina.

Agradecimientos

La investigación desarrollada ha recibido financiación de dos proyectos: "Development of social robots to help seniors with cognitive impairment" (ROBSEN), financiado por el Ministerio de Economía y Competitividad; y "RoboCity2030-III-CM", financiado por la Comunidad de Madrid y cofinanciado por los Fondos Estructurales de la Unión Europea.

Referencias

- [1] F. H. Wilhelm, A. S. Kochar, W. T. Roth, and J. J. Gross, "Social anxiety and response to touch: incongruence between self-evaluative and physiological reactions," *Biological Psychology*, vol. 58, no. 3, pp. 181–202, 2001.
- [2] K. Altun and K. E. MacLean, "Recognizing affect in human touch of a robot," *Pattern Recognition Letters*, vol. 66, pp. 31–40, nov 2015.
- [3] B. D. Argall and A. G. Billard, "A survey of tactile human-robot interactions," *Robotics and Autonomous Systems*, vol. 58, pp. 1159–1176, October 2010.
- [4] W. D. Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf, "Design of a therapeutic robotic companion for relational, affective touch," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pp. 408–415, Aug 2005.
- [5] D. Silvera-Tawil, D. Rye, and M. Velonaki, "Interpretation of social touch on an artificial arm covered with an EIT-based sensitive skin," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 489–505, 2014.
- [6] D. Silvera Tawil, D. Rye, and M. Velonaki, "Touch modality interpretation for an EIT-based sensitive skin," in *2011 IEEE International Conference on Robotics and Automation*, pp. 3770–3776, IEEE, may 2011.
- [7] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," tech. rep., Stanford University, 1998.
- [8] M. D. Cooney, S. Nishio, and H. Ishiguro, "Recognizing affection for a touch-based interaction with a humanoid robot," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1420–1427, IEEE, oct 2012.
- [9] Y. Ohmura, Y. Kuniyoshi, and A. Nagakubo, "Conformable and scalable tactile sensor skin for curved surfaces," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pp. 1348–1353, IEEE, 2006.
- [10] J. Paradiso and N. Checka, "Passive acoustic sensing for tracking knocks atop large interactive displays," in *Proceedings of IEEE Sensors*, vol. 1, pp. 521–527, IEEE, 2002.
- [11] C. Harrison and S. E. Hudson, "Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces," in *Proceedings of the 21st annual ACM symposium on User*

- interface software and technology - UIST '08*, (New York, New York, USA), p. 205, ACM Press, October 2008.
- [12] R. Murray-Smith, J. Williamson, S. Hughes, and T. Quaade, “Stane: synthesized surfaces for tactile input,” in *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, (New York, New York, USA), p. 1299, ACM Press, apr 2008.
- [13] P. Lopes, R. Jota, and J. A. Jorge, “Augmenting touch interaction through acoustic sensing,” in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11*, (New York, New York, USA), p. 53, ACM Press, nov 2011.
- [14] M. A. Salichs, R. Barber, A. M. Khamis, M. Malfaz, J. F. Gorostiza, R. Pacheco, R. Rivas, A. Corrales, E. Delgado, and D. Garcia, “Maggie: A robotic platform for human-robot social interaction,” in *2006 IEEE Conference on Robotics, Automation and Mechatronics*, pp. 1–7, June 2006.
- [15] G. Wang, P. R. Cook, and S. Salazar, “Chuck: A strongly-timed computer music language,” *Computer Music Journal*, vol. 39, no. 4, pp. 10–29, 2015.
- [16] F. Alonso-Martin, M. Malfaz, J. Sequeira, J. Gorostiza, and M. A. Salichs, “A Multimodal Emotion Detection System during Human-Robot Interaction,” *Sensors*, vol. 13, no. 11, pp. 15549–15581, 2013.
- [17] A. Klapuri, *Introduction to Music Transcription*, p. 8. Springer.
- [18] A. Cheveign and H. Kawahara, “Yin, a fundamental frequency estimator for speech and music,” *J. Acoust. Soc. Am.*, vol. 111, pp. 1917–1930, 2002.
- [19] P. McLeod and G. Wyvill, “A smarter way to find pitch,” in *Proceedings of International Computer Music Conference (ICMC)*, (Barcelona, Spain), September 2005.
- [20] E. Larson and R. Maddox, “Real-time time-domain pitch tracking using wavelets,” in *Proceedings of the University of Illinois at Urbana Champaign Research Experience for Undergraduates Program*, (Champaign, IL, USA), August 2005.
- [21] T. K. Ho, “Random decision forest,” in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pp. 278–282, August 1995.
- [22] D. Aha and D. Kibler, “Instance-based learning algorithms,” *Machine Learning*, vol. 6, pp. 37–66, 1991.
- [23] J. Platt, “Sequential minimal optimization: A fast algorithm for training support vector machines,” tech. rep., Microsoft Research, April 1998.