

EVALITA 2011



NXT Speech

**Utilizzo di comandi vocali nell'interazione con robot
Lego Mindstorm NXT**

CANGIANO GENNARO

Università degli Studi di Napoli Federico II



Indice generale

1.Introduzione.....	3
2.Hardware	3
2.1.Lego Mindstorm NXT.....	3
2.2.Smartphones.....	4
3.Software.....	4
3.1.NXT Speech.....	4
4.Esempio di grammatica.....	8

1. Introduzione

Il seguente documento ha lo scopo di descrivere l'utilizzo del riconoscitore embedded Loquendo ASR per dispositivi mobili (smartphone con sistema operativo Google Android) nella interazione con robot Lego Mindstorm NXT. La progettazione e l'utilizzo di robot NXT, per problematiche relative alla interazione di questi in ambienti appositamente riprodotti, esulano dalla trattazione di questo documento, basti sapere che i dispositivi in questione sono stati programmati per svolgere un determinato compito. L'integrazione di una "componente vocale" può essere vista come uno strumento fornito all'utente umano per intervenire nel compito assegnato al robot, nel caso in cui questo si trovi in situazioni di stallo.

2. Hardware

2.1. Lego Mindstorm NXT



Per Lego Mindstorm NXT si intende un robot-kit commerciale dell'azienda danese Lego messo sul mercato nel 2006 come evoluzione di un progetto precedente (RCX) iniziato sin dagli ultimi anni '90. La particolarità del corredo fornito con una tipica confezione del prodotto commerciale sta nel comprendere oltre i classici

mattoncini, anche folta serie di nuovi meccanismi hardware conosciuti come sensori e servomotori e una nuova tipologia di mattoni più grandi, detti programmabili (gli NXT brick), che assurgono alla funzione di componenti-cervello del sistema. Si tratta di una vera e propria serie di piccoli computer programmabili dotati di svariate connessioni a cui è possibile connettere un certo numero di sensori e servomotori. Ciascuno degli NXT brick è infatti una macchina in grado di ricevere l'input da un numero massimo di sensori pari a 4 ed è in grado di controllare fino a 3 servomotori attraverso l'ausilio di cavi di tipologia RJ12. L'NXT, oltre ad un display LCD monocromatico con matrice di 100x64 pixel, dispone di un altoparlante che può riprodurre file sonori campionati a 8 kHz, un hardware bluetooth per la comunicazione con altri NXT, personal computer e simili. La corrente è fornita nella versione per utenti normali da 6 batterie AA (1.5 V ognuna) mentre la versione "educational" da una batteria ricaricabile Li-Ion e il relativo carica batterie.

Da notare che la capacità di elaborazione del brick è estremamente limitata, così come le altre componenti hardware che costituiscono i vari sensori, che risultano dunque di qualità non eccelsa.

2.2. Smartphones

I dispositivi mobili sono oggi giorno diffusissimi grazie alle loro sempre più potenti capacità di calcolo, di memorizzazione e all'integrazione di hardware di buon livello: fotocamere, dispositivi bluetooth, gps, wi-fi.

Soprattutto negli ultimi anni vi è stato un rapido diffondersi di smartphone con veri e propri sistemi operativi, famiglia di cui fanno parte i famosissimi prodotti equipaggiati con Google Android. Fama dovuta al carattere open source del sistema ed alle ampie possibilità di sviluppo messe a disposizione dalla stessa Google Inc.: chiunque può sviluppare applicazioni per smartphone attraverso l'utilizzo del software development kit disponibile online, un framework di programmazione java.

3. Software

Il punto di incontro tra le due tipologie di dispositivi sopra elencati è proprio la presenza di un possibile canale di comunicazione offerto dai rispettivi hardware bluetooth. Inoltre, dal punto di vista software, è essenziale la possibilità di utilizzo del Loquendo ASR da codice scritto nel linguaggio JAVA, lo stesso utilizzato dai dispositivi con sistema Android. Per quanto riguarda i robot NXT, la versione retail del kit include software per scrivere programmi che possono essere eseguiti su PC e Macintosh. Il software in dotazione è basato su LabVIEW di National Instruments e fornisce un linguaggio di programmazione visuale per scrivere semplici programmi e per farne il download sul dispositivo. Tuttavia è possibile l'utilizzo di leJOS NXJ, un ambiente di sviluppo Java introdotto di recente per la programmazione dei LEGO MINDSTORM NXT. LeJOS inoltre include un firmware sostitutivo per gli NXT con incorporata una JVM, contiene librerie per la creazione del file binario caricabile ed eseguibile sull'NXT ed altre migliorie per i programmatori. In ultimo, ma non per importanza, leJOS è un progetto open source portato avanti dal portale SourceForge.

3.1. NXT Speech

Come sarà stato possibile intuire, l'introduzione di un approccio vocale per l'interazione con i robot NXT è stato trattato in un progetto già in corso d'opera che richiedeva una interazione completamente autonoma tra più robot. L'utilizzo introdotto è stato dunque limitato ad azioni da far compiere ad un robot, del tipo: girare a destra o a sinistra di un certo numero di gradi, andare avanti od indietro di un certo numero di centimetri, chiamare un altro robot

(nel senso di far giungere, nei pressi della posizione occupata, un altro robot), il ritorno alla posizione di partenza.

L'interazione tra l'utente umano ed il robot avviene, dunque, seguendo lo schema rappresentato in figura:



L'utente pronuncia una frase (comando) in prossimità dello smartphone, il riconoscitore Loquendo ASR interpreta l'audio ricevuto e restituisce una stringa di testo che rappresenta la trascrizione di ciò che è stato pronunciato.

Il passo successivo è quello di interpretare il comando, ossia analizzare il testo che l'ASR ha restituito e tradurlo in comandi da inviare al robot NXT. Per fare ciò, viene effettuato un parsing sintattico per la ricerca di parole chiave (“concetto”) all'interno della stringa di testo: ogni concetto è un parola fondamentale ed è riconducibile ad una delle seguenti classi concettuali:

- IdRobot: rappresenta il soggetto o l'oggetto della frase pronunciata
- NxtActions: costituisce l'azione da compiere
- NxtDisplacements: in che misura dell'azione deve essere compiuta

Per non rendere complessa la struttura del codice degli NXT e per limitazioni relative al protocollo bluetooth, si è scelto di comunicare con uno solo dei robot, questo ha reso possibile scindere in due categorie le possibili frasi da pronunciare:

- <SOGGETTO> <AZIONE> <SPOSTAMENTO>
- <AZIONE> <OGGETTO>

Nel primo caso, si tratta di azioni che solo il robot ricevente (numero 1) può eseguire; nel secondo caso, l'azione viene implicitamente comunicata al robot numero 1, ma vi è un oggetto di tale azione, ossia l'altro robot coinvolto nella comunicazione al fine di coinvolgerlo nelle operazioni.

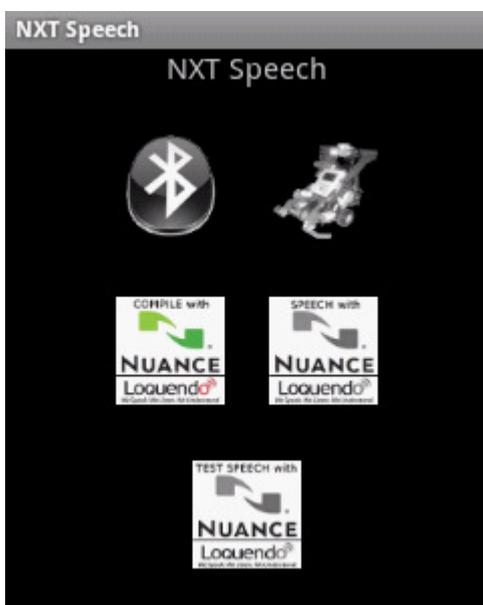
In definitiva, dunque, le possibili combinazioni dei concetti pronunciabili dall'operatore in relazione alle classi concettuali sopra elencate, possono essere così riassunte:

- <IdRobot> <NxtActions> <NxtDisplacements>
- <NxtActions> <IdRobot>

Esempi di frasi pronunciabili appartenente alle categorie sopra elencate sono:

- ROBOT NUMERO 1 VAI AVANTI DI QUARANTA CENTIMETRI
- CHIAMA IL ROBOT NUMERO DUE

L'applicazione NXT Speech si presenta nelle seguenti interfacce grafiche:



Nell'immagine a sinistra viene riportata l'applicazione al primo avvio, le icone sono quasi tutte monocromatiche poiché l'utilizzo corretto prevede dei passi fondamentali come spiegato di seguito.



Attivazione del bluetooth: è obbligatoriamente il primo passo da compiere nel caso si voglia utilizzare l'applicazione per il fine pratico descritto in questo documento (cioè comunicare comandi al robot NXT).



Accoppiamento col robot NXT desiderato: è un altro passo fondamentale per comunicare con il robot. L'opzione è disponibile subito dopo l'attivazione del dispositivo bluetooth: all'utente viene mostrato l'elenco dei robot già accoppiati con lo smartphone oppure è possibile effettuare una scansione per registrarne altri.



Compilazione della grammatica: è possibile compilare la grammatica fornita con l'applicazione indipendentemente dall'accoppiamento del dispositivo. Ovviamente, se il robot è stato accoppiato allo smartphone è possibile comunicargli i comandi; in caso contrario, l'unica possibilità di utilizzo dell'applicazione è quella di testare il riconoscimento della frase pronunciata (opzione introdotta solo a scopo dimostrativo, vedi ultima opzione).



Riconoscimento vocale: una volta effettuati tutti e tre i passi precedenti, è possibile utilizzare l'applicazione facendogli riconoscere la frase ed inviando, in automatico, al robot NXT il comando pronunciato.



Test del riconoscimento: opzione disponibile una volta compilata la grammatica ed utile, nel caso non si disponga di un robot NXT, solo per testare il riconoscitore Loquendo ASR.

Al seguente link è disponibile per il download un video dimostrativo dell'applicazione in funzione:

http://dl.dropbox.com/u/16316582/NXTspeech_Evalita2011_Cangiano.mp4

4. Esempio di grammatica

```
#ABNF 1.0 UTF-8;
language it-IT;
mode voice;
tag-format <semantics\1.0>;
root $Speech;

public $Speech = [$Varie] [$IdRobot] $NxtActions [$IdRobot] [$Varie];

private $IdRobot =
    esploratore | ( ( ( [il] robot ) | ( [il] figlio ) | nxt ) [numero] $Number );

private $Number = uno | due | tre | primo | secondo | terzo;

private $NxtActions =
    ( ( gira | svolta ) $Verso [di] $Gradi [gradi] ) |
    ( ( ( [vai] avanti ) | avanza ) [di] $Centimetri [centimetri] ) |
    ( ( ( [vai] indietro ) | indietreggia | arretra ) [di] $Centimetri
[centimetri] ) |
    ( manda [ a casa ] ) | chiama | ( ( torna | ritorna ) [ ( a casa ) | ( alla tana )
]);

private $Verso = ( [a | verso] ( destra | sinistra ) );

private $Gradi =
    dieci |
    quindici |
    trenta |
    quarantacinque |
    sessanta |
    settantacinque |
    novanta |
    centottanta;

private $Centimetri =
    venti |
    quaranta |
    ottanta;

private $Varie =
    prego |
    ( per favore ) |
    grazie;
```