

Seminario di cultura digitale
**Domotica, Interaction Design e Riconoscimento
Vocale**

Nicola Leonardi MAT. 566547

Informatica Umanistica

Anno accademico 2017/2018

Tabella dei contenuti

1	Introduzione.....	3
1.1	<i>La domotica una disciplina multisetoriale.....</i>	10
1.1.1	Informatica	10
1.1.2	Visual Design	11
1.1.3	Usabilità e User Experience Design.....	12
1.1.4	Architettura dell'informazione.....	15
1.1.5	Ergonomia.....	16
2	Le principali tecnologie.....	17
2.1	<i>Le tecnologie professionali</i>	17
2.2	<i>Le tecnologie open-source</i>	22
2.2.1	Arduino	22
2.3	<i>Le tecnologie plug-and-play.....</i>	23
2.3.1	Apple Home Kit	23
2.3.2	Google Home	25
2.3.3	IFTTT.....	26
3	Il controllo vocale	27
3.1	<i>Multimodalità</i>	27
3.2	<i>Voice User Interface (VUI).....</i>	28
3.3	<i>Violazione dei principi di usabilità e di design</i>	29
3.4	<i>Una tappa inevitabile?.....</i>	30
3.5	<i>Controllo vocale: un ambito di studio ancora aperto</i>	31
Parte II: Domotica e controllo vocale		
4	CMUSphinx: descrizione architettura	31

4.1	<i>Analisi del modello attuale per italiano</i>	35
4.2	<i>Descrizione Philips Hue</i>	36
4.3	<i>Creazione Language Model ad hoc</i>	37
4.4	<i>Verifica funzionamento</i>	38
5	Commenti conclusivi	38
6	Bibliografia	39

Abstract

L'elaborato si articolerà su un percorso che partendo da un overview sullo stato della domotica ai giorni nostri toccherà nozioni di interaction design e multimodalità e dopo una breve descrizione dei componenti e delle modalità di utilizzo del toolkit CMUSphinx, descriverà un'applicazione pratica di controllo vocale integrato con il sistema di controllo luci Philips HUE.

Keywords: Domotica, Usabilità, Interaction Design, multimodalità, riconoscimento vocale, CMUSphinx, Crestron, AMX, Control4, Philips HUE

Tempo stimato di lettura: 20 minuti.

1 Introduzione

La *domotica*, dal francese domotique, parola composta dal latino domus 'casa' e (informa)tique, è la *scienza interdisciplinare* che si occupa dell'impiego delle tecnologie informatiche per il miglioramento della qualità della vita nella casa e più in generale negli ambienti civili.

La domotica (chiamata anche home-automation) è spesso associata al concetto di "casa intelligente" ovvero un ambiente, opportunamente progettato e tecnologicamente attrezzato, dove la tecnologia si integra al "tradizionale", dove apparati e sistemi sono in grado di svolgere funzioni "intelligenti" in parziale autonomia coordinandosi fra loro (ad esempio un parametro ambientale diventa un input per il sistema che risponde proponendo un certo output); tali funzionalità vengono *programmate* a priori in funzione delle necessità dall'utente o più recentemente possono essere gestite in totale autonomia *adattandosi* alle abitudini dell'utente secondo algoritmi di apprendimento automatico, approcci ottimali per il raggiungimento della massima personalizzazione ed efficienza.

Ad un livello superiore si parla di building-automation (automazione degli edifici), ovvero un "edificio intelligente" che grazie ad una progettazione tecnica adeguata riesce ad integrare in maniera efficiente e coordinata i diversi impianti tecnologici presenti nel complesso architettonico come ad esempio impianto di illuminazione, di riscaldamento, sistema di sicurezza ed anti-intrusione, irrigazione, intrattenimento audio-visivo. Ricadono in questo ambito aeroporti, chiese, ospedali, scuole, piazze, cinema, teatri, centri benessere, ecc.

Tra i benefici che l'adozione di un sistema di domotica può comportare si possono elencare:

- miglioramento della qualità della vita espressa anche in termini di personalizzazione
- maggiore sicurezza (fumi, gas) e controllo
- facilità di utilizzo tramite gestione integrata delle nuove tecnologie (eliminazione dei vari telecomandi per tv, decoder, condizionatore, cancello, ecc).
- evoluzione temporale dell'impianto alle nuove esigenze (l'impianto non è "sculpto nella pietra" ma programmabile nel tempo) -> riduzione dei costi di gestione.
- risparmio energetico (gestione ottimizzata delle risorse)

Uno degli aspetti fondamentali apportati dalla domotica è quello della *semplificazione* e della *personalizzazione* dei controlli, tematica che può essere associata al concetto di *accessibilità*, ovvero l'insieme di tecniche per rendere fruibile l'impianto a qualsiasi persona indipendentemente dalle capacità psico-fisiche della stessa. (es. anziani o malati).

Come vedremo la costruzione di un'interfaccia *user-friendly*, ovvero che rispetti i canoni di usabilità universalmente riconosciuti, è uno degli aspetti cardini del progetto domotico.

Spesso la domotica viene intesa come *intrattenimento* quando l'esperienza domotica si concentra sulla gestione dell'impianto audio-visivo e ludico-tecnologico, ma questo ambito è solamente un sottoinsieme delle discipline e delle tematiche che la domotica abbraccia.

Un elenco (sicuramente non esaustivo) delle aree di interesse della domotica è quello che segue:

- illuminazione
- distribuzione elettrica
- irrigazione
- ambienti esterni (piscina, coperture da giardino)
- climatizzazione e riscaldamento
- telefonia VOIP

- gestione accessi, antifurto, videosorveglianza
- audio-video
- tende e tapparelle

Il grafico che segue mostra il mercato americano domotico rappresentato per tipologia nel corso degli anni con stime per il futuro (fonte <http://www.grandviewresearch.com/industry-analysis/smart-homes-industry>).

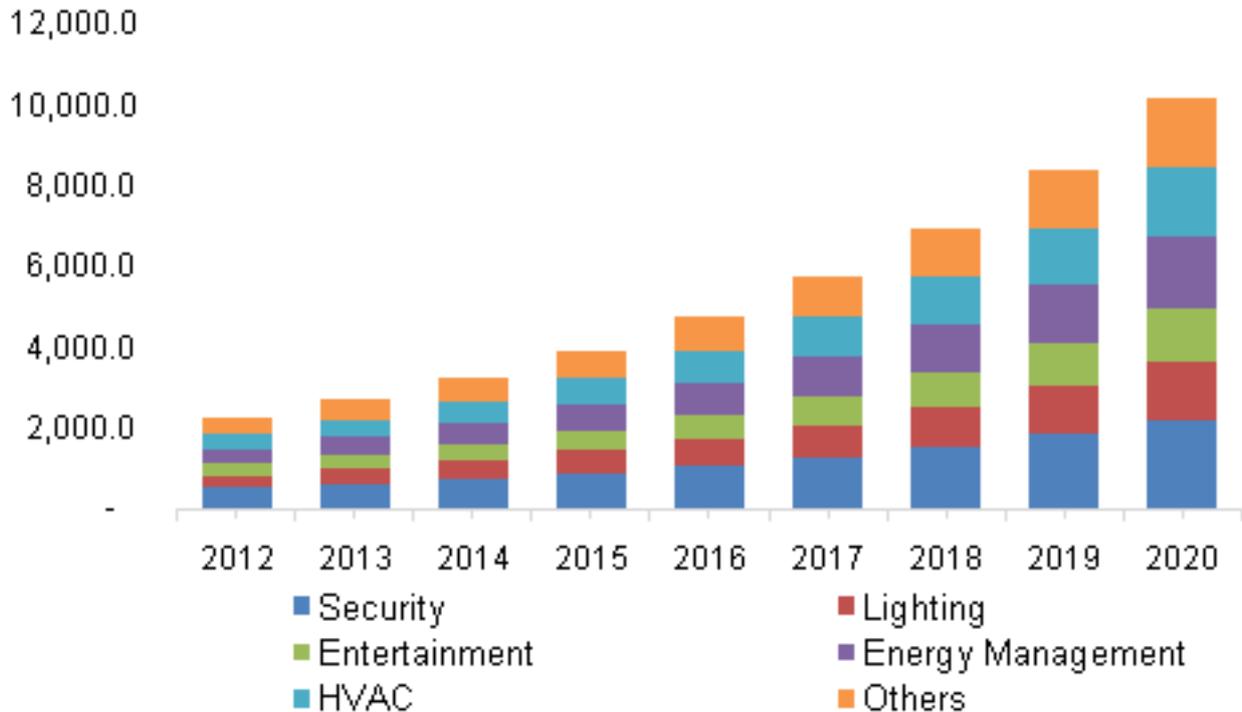


Figura 1. Tipologie e distribuzione di installazioni domotiche

Risulta chiaro come il mercato stia aumentando e con esso l'eterogeneità di tipologia di installazioni ed ambiti coinvolti.

Recentemente alcuni produttori (es. Samsung) hanno aperto molti apparati domestici al controllo e supervisione tramite opportune API (Application Programming Interface). Sono dunque entrati nel campo della domotica anche dispositivi comuni come frigoriferi, lavastoviglie, forni, ecc.

La figura che segue mostra l'interfaccia grafica di un sistema integrato per il controllo di vari aspetti e funzionalità in un caratteristico ambiente domotico.



Figura 2. Esempio di interfaccia di controllo domotico

Come vedremo più avanti nella sezione dedicata alla descrizione delle varie tecnologie, ci sono tre elementi chiave per un sistema domotico:

- le interfacce di input
- il controllore
- le interfacce di output

Le interfacce di input rappresentano tutto ciò che può scatenare un evento ritenuto di interesse al fine del progetto.

Fanno parte di questa categoria elementi come i semplici bottoni, pulsantieri e telecomandi, ma anche sensori, touch screen dedicati, input da riconoscimento vocale (a cui sarà dedicata una specifica descrizione) e gesture.

Il controllore è l'unità informatica intelligente (o l'insieme delle unità distribuite come nel caso [KNX](#)) che ricevendo "gli stimoli" dalle interfacce di input, le processa, elabora e coordina al fine di generare un output coerente.

L'*output* può essere di *tipo funzionale* (cambio di stato, attivazione-disattivazione, on-off, acceso-spento) o un feedback rappresentato sul touch-screen, su un led sulla pulsantiera, un segnale sonoro ecc. oppure un *recupero di un'informazione* (es. di stato, di evento, di log).

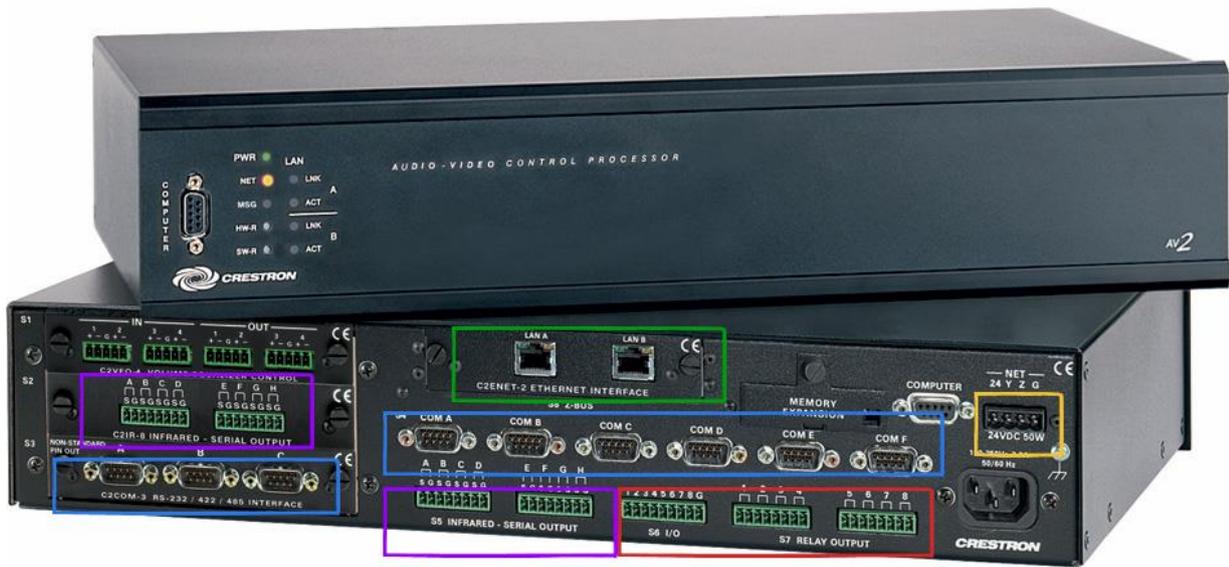


Figura 3. Controllore e tipologie di IN/OUT

I tre diversi gruppi di elementi sono *interconnessi* attraverso varie tipologie di connessione (rete LAN, porte seriali, onde convogliate, bus dedicati, onde radio) e dialogano tra loro grazie “all’intelligenza” sul controllore, che grazie alla fase di programmazione attuata dal *system integrator* riesce a far dialogare sistemi e tecnologie eterogenee, introducendo un livello comune di dialogo altrimenti inesistente.

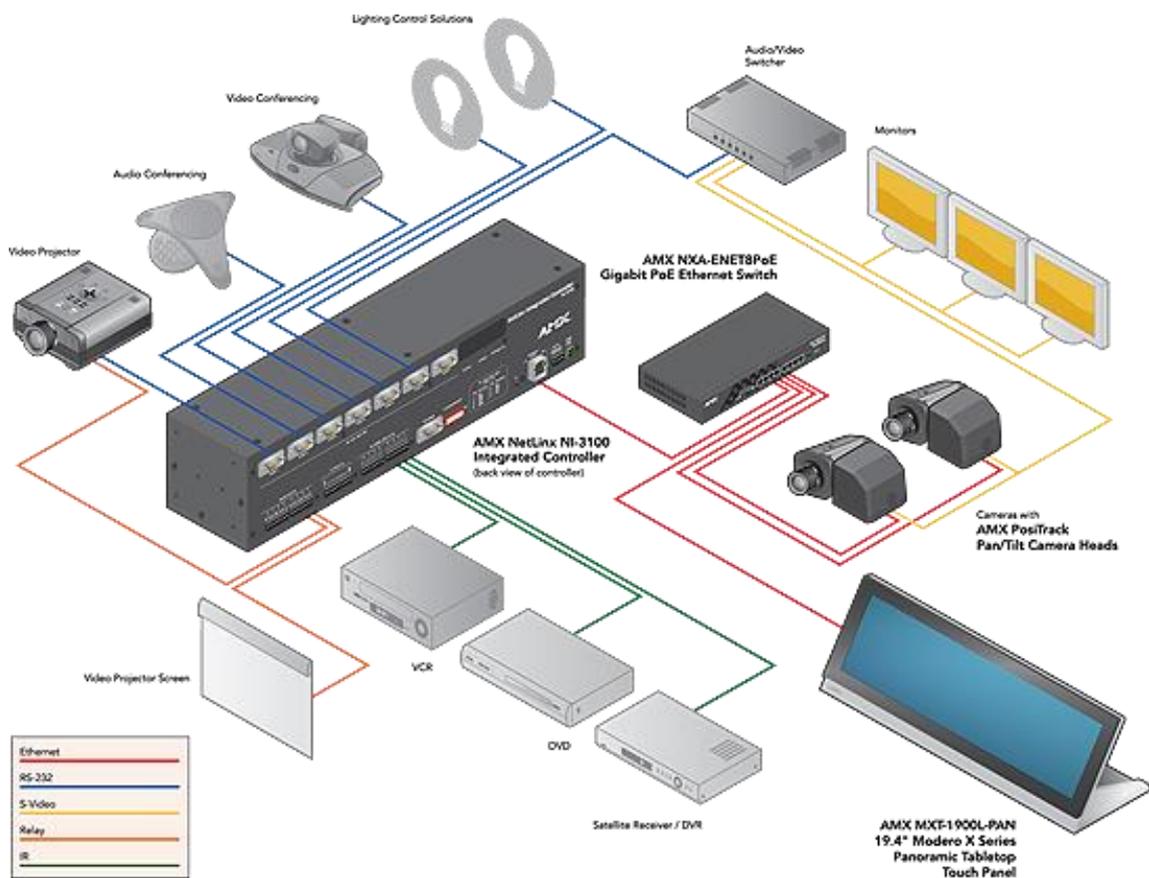


Figura 4. Esempio di installazione domotica e di interconnessione fra dispositivi (www.amx.com)

Basics of Control Review

- Match each property to the appropriate serial control type: (some properties have more than one answer)

RS-232 RS-422 RS-485 TCP/IP

- Uses a balanced signal to reduce interference
- Maximum run length of 50 feet
- Only multi-drop
- Single-ended or multi-drop
- Connects only one transmitter and receiver
- Maximum run length of 4000 feet
- Uses twisted pair (Cat5) wiring

Figura 5. Un tipico overview delle modalità di comunicazione

Il system integrator lavora in relazione alle API (Application Programming Interface) che il produttore del dispositivo rende accessibili eventualmente previa creazione di un account di tipo developer.

1. Lights API

1.1. Get all lights

URL	/api/<username>/lights
Method	GET
Version	1.0
Permission	Whitelist

1.1.3. Response example

```
{
  "1": {
    "state": {
      "on": true,
      "bri": 144,
      "hue": 13088,
      "sat": 212,
      "xy": [0.5128,0.4147],
      "ct": 467,
      "alert": "none",
      "effect": "none",
      "colormode": "xy",
      "reachable": true
    },
    "type": "Extended color light",
    "name": "Hue Lamp 1",
    "modelid": "LCT001",
    "swversion": "66009461",
    "pointsymbol": {
      "1": "none",
      "2": "none",
      "3": "none",
      "4": "none",
      "5": "none",
      "6": "none",
      "7": "none",
      "8": "none"
    }
  },
  "2": {
    "state": {
      "on": false,
      "bri": 0,
      "hue": 0,
      "sat": 0,
      "xy": [0,0],
      "ct": 0,
      "alert": "none",
      "effect": "none",
      "colormode": "hs",
      "reachable": true
    }
  },
}
```

Figura 6. API Philips HUE

L'aggiornamento continuo sulle ultime tecnologie e standard è un requisito fondamentale per questa tipologia di lavoro; molto spesso la situazione si capovolge dovendo andare alla ricerca di "vecchi" protocolli di comunicazione seriale o la gestione di standard "industriali" ormai datati come ad esempio di tipo modbus.

1.1 La domotica: una disciplina multisetoriale

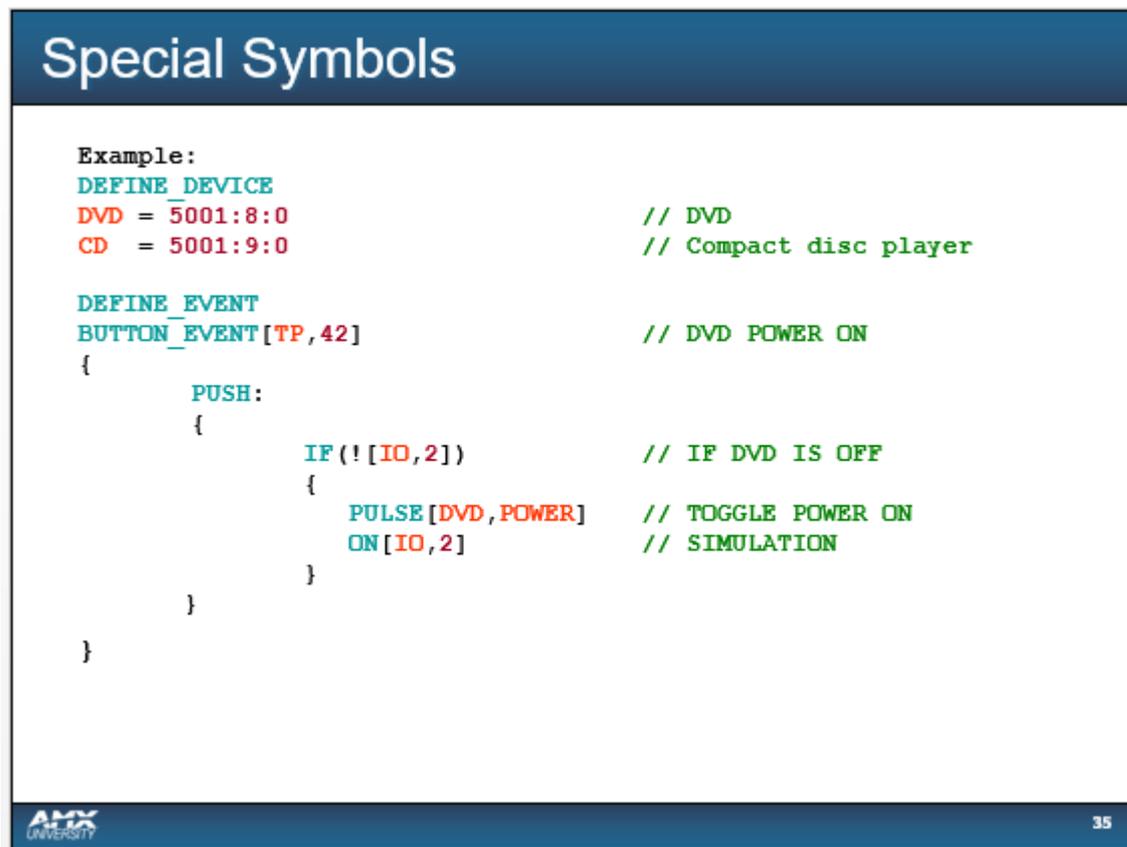
Come descritto nel paragrafo introduttivo risulta chiaro che la domotica rappresenta una disciplina multisetoriale che abbraccia molteplici ambiti sia scientifici sia umanistici.

La figura del *system integrator* dovrà quindi sviluppare l'abilità di districarsi su diversi ambiti tra i quali:

- Informatica
- Visual Design
- Usabilità / User Experience
- Architettura dell'informazione
- Ergonomia

1.1.1 Informatica

Come accennato non esiste un progetto domotico che escluda conoscenze di programmazione multilinguaggio (linguaggi proprietari e comuni del tipo HTML, Javascript), di database (query SQL), di reti informatiche e di telecomunicazioni.



```
Special Symbols

Example:
DEFINE_DEVICE
DVD = 5001:8:0           // DVD
CD = 5001:9:0           // Compact disc player

DEFINE_EVENT
BUTTON_EVENT[TP,42]     // DVD POWER ON
{
    PUSH:
    {
        IF(![IO,2])     // IF DVD IS OFF
        {
            PULSE[DVD,POWER] // TOGGLE POWER ON
            ON[IO,2]       // SIMULATION
        }
    }
}
```

Figura 7. Esempio di programmazione in linguaggio proprietario AMX

Per certi progetti è sufficiente un livello conoscitivo da configuratore, ma per certi tipologie di impianti è necessaria una conoscenza approfondita di vari aspetti (es reti private VPN, ottimizzazione di query SQL, ecc).

1.1.2 Visual Design

Gli input/output sono prevalentemente gestiti a livello visivo, dunque, la progettazione grafica, la prototipazione e il concept design sono processi naturali ed imprescindibili nella costruzione di un'interfaccia attrattiva.

Una caratteristica peculiare dell'interfaccia domotica è che spesso è un'interfaccia ad hoc sviluppata su chiare preferenze del committente. Quindi, mentre nel costruire un'interfaccia web rivolta ai molti si cercano di seguire i "trends del momento" (anche per una questione di "confidenzialità" dell'utente finale), nell'ambito domotico spesso ci si concentra (e spesso ci si scontra) sui gusti e peculiarità specifiche del singolo che spesso possono essere anche lontani dai canoni progettuali e visuali del momento.

La definizione delle icone, degli sfondi, delle proporzioni, dei punti focali ecc. sono tutti aspetti importantissimi per questa tipologia di progetti.



Figura 8. Design di icone

Da tener presente è l'ambiente ed il contesto finale in cui l'interfaccia si collocherà; un casa "dell'Ottocento" potrebbe richiedere uno stile completamente diversa da un loft moderno, un centro benessere "tematico" (es natura) uno completamente diverso ancora e così via.

Aspetti come "*adaptive/responsive interface*" sono all'ordine del giorno dato che saranno molteplici i dispositivi con cui si vorrà gestire l'impianto (smatphone, tablet, smart tv); "*end user interface*" è un altro concetto chiave dato che per esempio ad un bambino potranno essere limitate le funzionalità di controllo rispetto ad un adulto (ad esempio inibita la gestione dell'antiintrusione o dell'irrigazione).

Un altro concetto cruciale è l'adattamento dell'interfaccia in base alla posizione dell'utilizzatore e alle condizioni fisico-ambientali dell'ambiente circostante: "*environment interface*". Ad esempio la tecnologia potrebbe rilevare la posizione dell'utente (tramite GPS se outdoor o per esempio tramite sensori dedicati RFID, IR, WLAN, bluetooth) e aggiornare l'interfaccia ai contenuti (o controlli) relativi a quella zona (caratteristica ad esempio utile alla fruizione dei contenuti audio-visivi in un museo).

La capacità di un'interfaccia di mutare in funzione dei parametri di tipo utente, tecnologia, ambiente ed aspetti sociali rientra in quello che viene definito **contesto d'uso**.

1.1.3 Usabilità e User Experience Design

La *User Experience* e l'usabilità sono entrambi indicatori di qualità di un'interfaccia o di un prodotto. Più precisamente, l'usabilità ci dice quanto tempo è necessario per imparare ad usare una cosa, con quanta efficienza la si usa poi, quanto si riesce a ricordarne il funzionamento, quanto è alta la probabilità di fare errori e la soddisfazione risultante quando la si usa.

La *User Experience* aggiunge a tutto questo i sentimenti. L'obbiettivo è il far sentire bene, immergere l'utente in un'esperienza gradevole e memorabile. Lo scopo è raggiungere la felicità.

Risulta palese come un progetto domotico debba metabolizzare questi concetti e prenderli in considerazione fin dalle prime fasi di sviluppo.

Tra le "regole" dell'usabilità ricordiamo le 10 euristiche di Jacob Nielsen:

- Visibilità dello stato del sistema
- Mapping col mondo reale
- Controllo e libertà
- Consistenza e standard
- Prevenzione dell'errore
- Favorire il riconoscimento piuttosto che il ricordo
- Flessibilità d'uso
- Design e estetica minimalista
- Aiutare l'utente a riconoscere, diagnosticare e recuperare l'errore
- Fornire aiuto e documentazione

ed i 7 principi di un buon design di Donald Norman [1]:

- Visibilità
- Feedback
- Affordance
- Significanti
- Vincolo
- Mapping
- Modello concettuale

Questi sono tutti concetti fondamentali nella progettazione di un'interfaccia domotica; verrà qua descritto solo il *mapping* rimandando al riferimento bibliografico per uno studio approfondito degli altri concetti .

Il *mapping* indica una relazione tra elementi di insiemi diversi.

Per restare sul semplice basta pensare al legame che esiste fra gli interruttori di una stanza e le rispettive luci.

Lo scopo ultimo di un buon mapping è rendere facile ed immediato la relazione fra comandi (ovvero le azioni) ed il risultato voluto evitando l'uso di etichette o simboli.

In generale quando per capire il funzionamento di un oggetto semplice servono delle etichette, delle icone o quant' altro siamo di fronte ad un esempio di cattivo design (Norman).

Il mapping spaziale ed il *mapping naturale*, ovvero sfruttare i riferimenti spaziali e fare riferimento al mondo reale, sono due approcci risolutivi per la progettazione dei comandi domotici.

Il posizionamento su una semplice piantina 2-D dei comandi per l'accensione delle luci ne è un classico esempio implementativo.

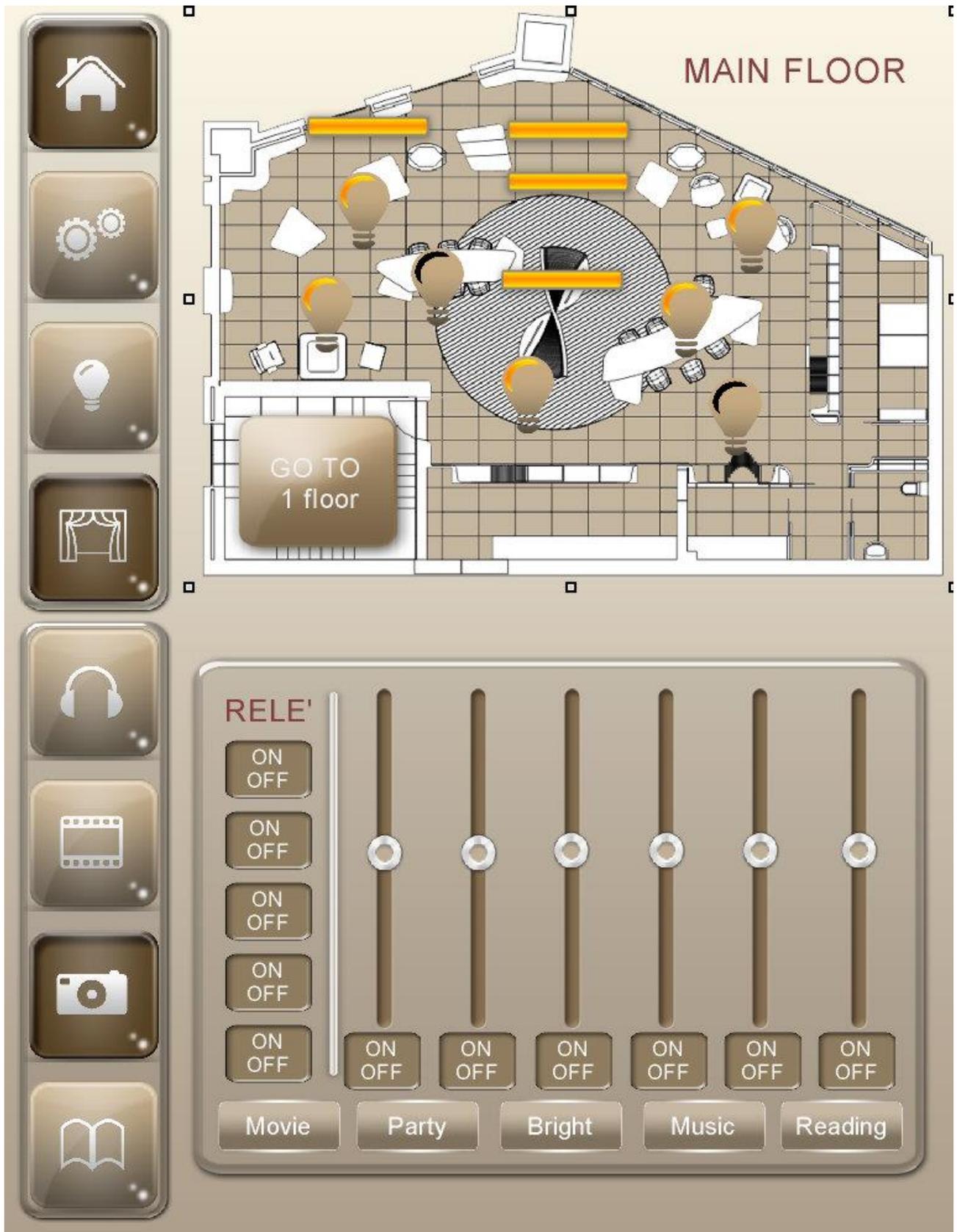


Figura 9. Mapping spaziale su un'interfaccia di controllo domotico

Dalla User Experience Design si recupera l'approccio circolare e ripetitivo della progettazione con un abbandono del classico modello a cascata:

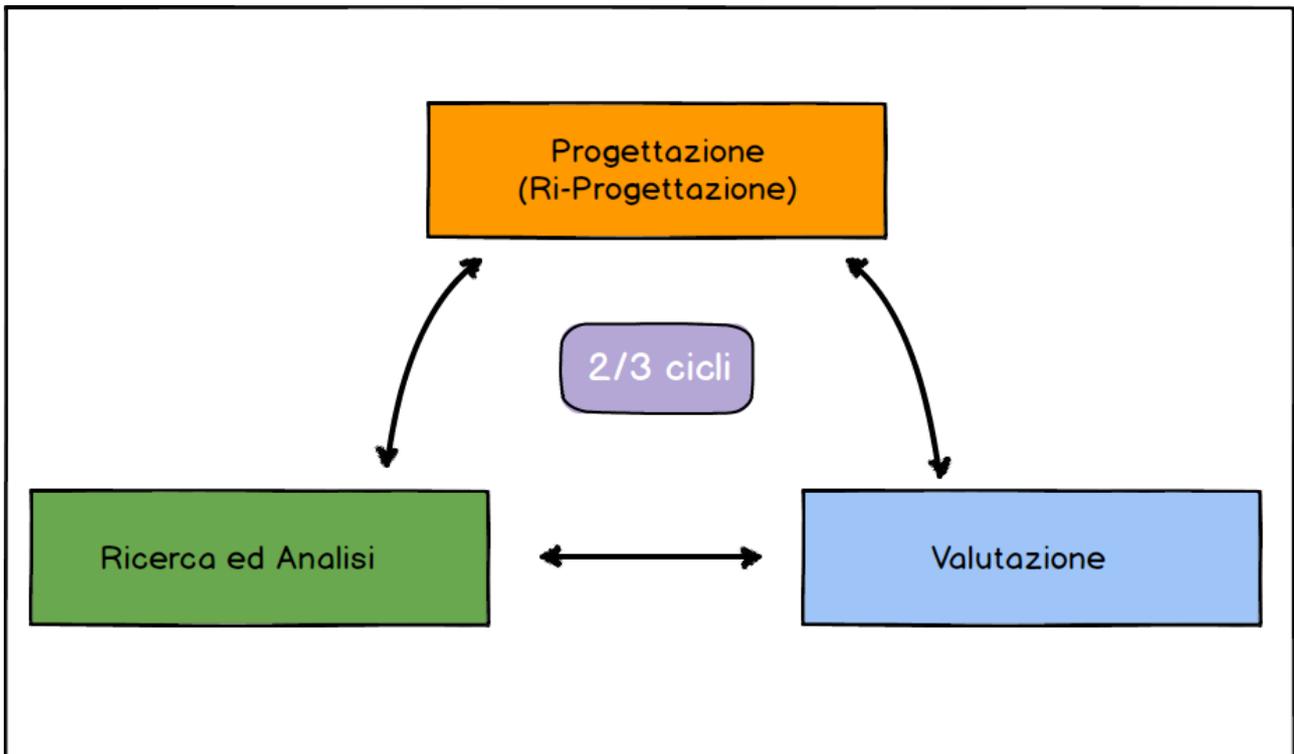


Figura 10. Fasi di elaborazione di un progetto domotico

Come accennato in precedenza l'ultimo ciclo sarà effettuato col committente finale e personalizzato alle sue richieste peculiari.

1.1.4 Architettura dell'informazione

Dagli studi sull'architettura dell'informazione (Information Architecture) si recuperano concetti e principi quali:

Organizzazione: dare un senso ai contenuti dell'interfaccia in maniera che siano facilmente raggiungibili dall'utilizzatore.

I vari tipi di organizzazione possono essere classificati in:

- Organizzazione alfabetica
- Organizzazione cronologica
- Organizzazione spaziale
- Organizzazione per argomento
- Organizzazione per compito
- Organizzazione per utenza

Navigazione: passare da una sezione all'altra della nostra interfaccia.

Essa svolge tre funzioni sostanziali ovvero risponde alle domande:

- Dove sono?

- Dove posso andare?
- Come posso raggiungere ciò che desidero?

Etichettatura: come comunichiamo l'informazione.

- L'etichetta deve suggerire
- L'etichetta deve essere predittiva (senza alcuno sforzo cognitivo so dove mi porterà)
- L'etichetta deve essere descrittiva e distintiva
- Le etichette devono essere sequenziali a livello logico (per sfruttare la legge di Hick sul tempo di scelta in funzione del numero di scelte).
- Le etichette devono essere semplici e di linguaggio comune

Questi principi, adottatissimi nella costruzione di pagine web, solo allo stesso modo pietre miliari per la progettazione di un'interfaccia domotica.

La loro violazione porta a grosse problematiche di gestione dell'interattività come vedremo in seguito nell'analisi del controllo vocale.

1.1.5 Ergonomia

L'ergonomia è quella scienza che si occupa dell'interazione tra gli elementi di un sistema (umani e d'altro tipo) e la funzione per cui vengono progettati.

Risulta immediato come questo aspetto sia alla base di un buon progetto di domotica.

Spazi, ambienti, posizioni, limitazioni sono tutti aspetti con cui la progettazione "dell'ambiente intelligente" dovrà confrontarsi ed integrarsi.



Figura 11. Tavolo multimediale all'interno di una showroom interattiva

2 Le principali tecnologie

In questa sezione saranno descritte le principali tecnologie utilizzate in ambito home-building automation. Sarà fatta una panoramica a partire da quelle professionali commercializzate in tutto il mondo per poi passare alle tecnologie "home-made" e le cosiddette tecnologie "plug-and-play" (nelle quali ricadranno anche le soluzioni *end-user-development*).

2.1 Le tecnologie professionali

Come tecnologie professionali verranno descritte quelle ritenute più rivelanti da un punto di vista "storico" e dal punto di vista del "peso" nel mercato di riferimento.

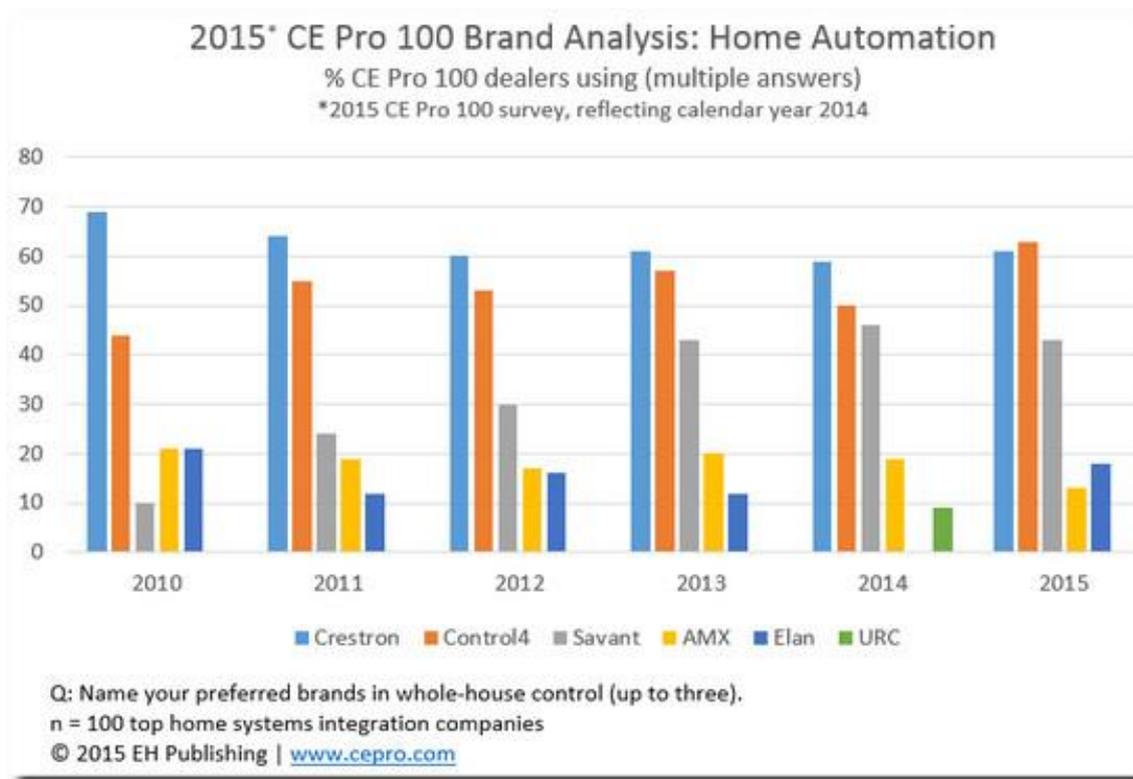


Figura 12. Il mercato delle principali tecnologie domotiche (www.cepro.com)

Come si può notare dal grafico (con indicatori a livello mondiale non europei) [Crestron](#), [AMX](#) e [Control 4](#) fanno parte del gruppo delle 5 tecnologie più utilizzate.

Il “valore” mostrato di AMX è un po’ fuorvincente, in quanto ci sono ambiti molto forti, come quello navale croceristico europeo, in cui AMX fa ancora, vuoi anche per motivi “storici”, da padrone.

Crestron, di origine europea, è molto forte su tutti i mercati (navale, civile, educativo) e anche in Italia è da considerarsi il leader del mercato.

Control4 è nata e diffusa intorno agli anni 2000 in USA (quasi 20 anni dopo rispetto a Crestron ed AMX) sta riscuotendo sempre molto più successo (soprattutto in USA) anche grazie alla semplicità di uso e programmazione (pagata a discapito della flessibilità) che la rendono una buona tecnologia per gli impianti entry-level.

Alla base di queste tecnologie vi sono i concetti di controllore, di dispositivi di input e di output interconnessi via bus proprietari (es. Cresnet per Crestron) o tecnologie standard (LAN, zig-bee).

Un’altro concetto è quello di separazione fra logica informatica, che va sul controllore e grafica che va sulle interfacce visive di output.

Spesso le due attività sono fatte da personale diverso (informatici per la prima e grafici e designer per la seconda) in quanto gestite da programmi diversi (Simple Windows e Vtproe per Crestron ad esempio) dove i due “mondi” sono separati (gli elementi di congiunzione sono i join analogici, digitali, seriali).

Anche a livello di programmazione vi è una distinzione in livelli di complessità crescente, dalla programmazione a blocchi logici dove le funzionalità sono preconfigurate e standardizzate alla programmazione in linguaggi proprietari simil C, C++ (per Crestron Simple+).

La prima modalità di programmazione (che potremmo interpretare come un approccio del *tipo end user development*) è adatta agli impianti di complessità moderata e permette una flessibilità limitata nonostante le numerose famiglie di “blocchi standard” presenti nei database del software.

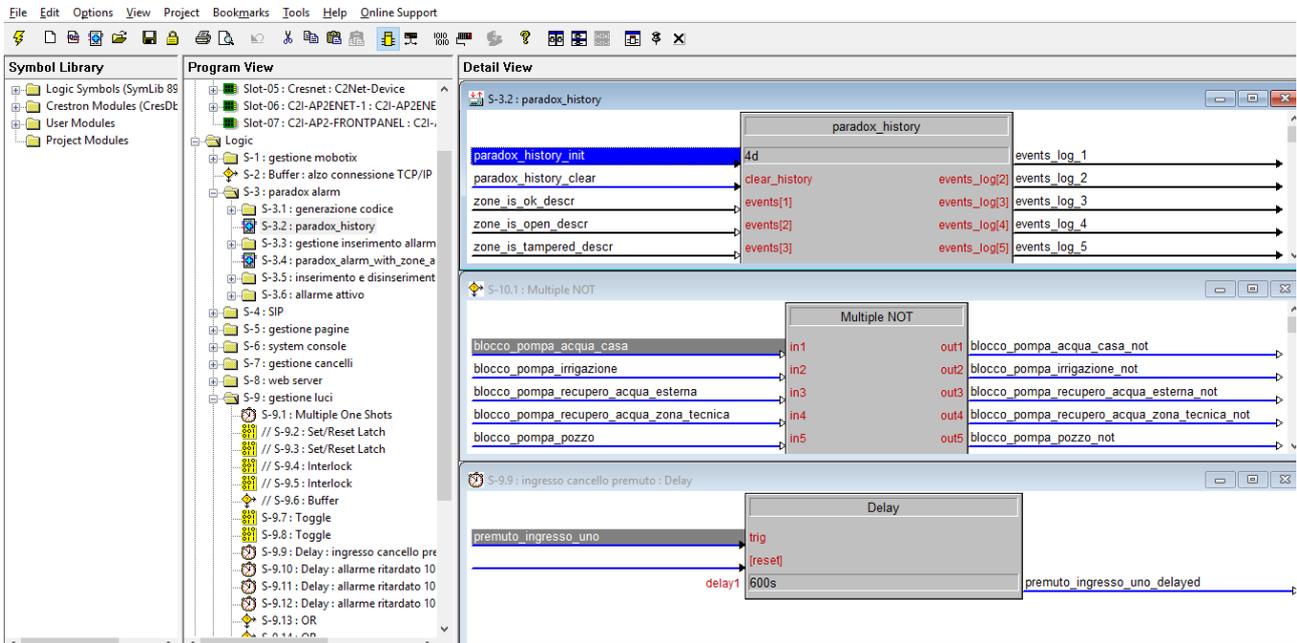


Figura 13. Programmazione a blocchi funzionali

Per impianti sofisticati, dove anche le performance hanno il loro peso, bisogna scendere al livello di programmazione pura.

```

337 FindClose();
338 EndFileOperations();
339 }
340 ELSE
341 {
342 nFileHandle = FileOpen ( FILENAME , _O_RDONLY );
343 If ( nFileHandle >= 0 )
344 {
345 For ( iCount = 1 to 4 )
346 {
347 ReadStructure( nFileHandle, S_ALARM[iCount], nTotalBytesRead );
348 If ( nTotalBytesRead < 0 )
349 {
350 #IF_DEFINED DEBUG
351 PRINT( "Error reading structure. Error code = %d\n", nTotalBytesRead );
352 #ENDIF
353 }
354 Else
355 {
356 A_Hour[iCount] = S_ALARM[iCount].s_A_Hour;
357 A_HourTemp[iCount] = S_ALARM[iCount].s_A_HourTemp;
358 A_Minute[iCount] = S_ALARM[iCount].s_A_Minute;
359 A_Days[iCount] = S_ALARM[iCount].s_A_Days;
360 AlarmTime[iCount] = S_ALARM[iCount].s_AlarmTime;
361 AMPM[iCount] = S_ALARM[iCount].s_AMPM;
362 #IF_DEFINED DEBUG
363 PRINT( "%d: hour = %d, hourtemp = %d, minutes = %d, alarmtime = %s, ampm = %s \n\n", iCount, A_Hour[iCount], A_Hc
364 #ENDIF
365 }
366 }
367 IF ( S_ALARM[1].s_A_Set = 1 )
368 ALARM_1_ACTIVE_FB = 1;
369 ELSE
370 ALARM_1_ACTIVE_FB = 0;
    
```

Figura 14. Programmazione a linee di codice

Tutte le tecnologie mettono a disposizione diverse tipologie di input/output dedicati (touch panels, tastierini, telecomandi, ecc.) e App sui diversi e-store.

Le tipologie di interconnessioni, la filosofia di funzionamento e gli elementi di input-output sono sostanzialmente gli stessi con piccole variazioni in funzione delle specifiche hardware/software della tecnologia abbracciata.

La figura che segue mostra un impianto con un controllore, degli extender e switcher di segnali audio/video, monitor e proiettori. La figura è specifica per un impianto Crestron.

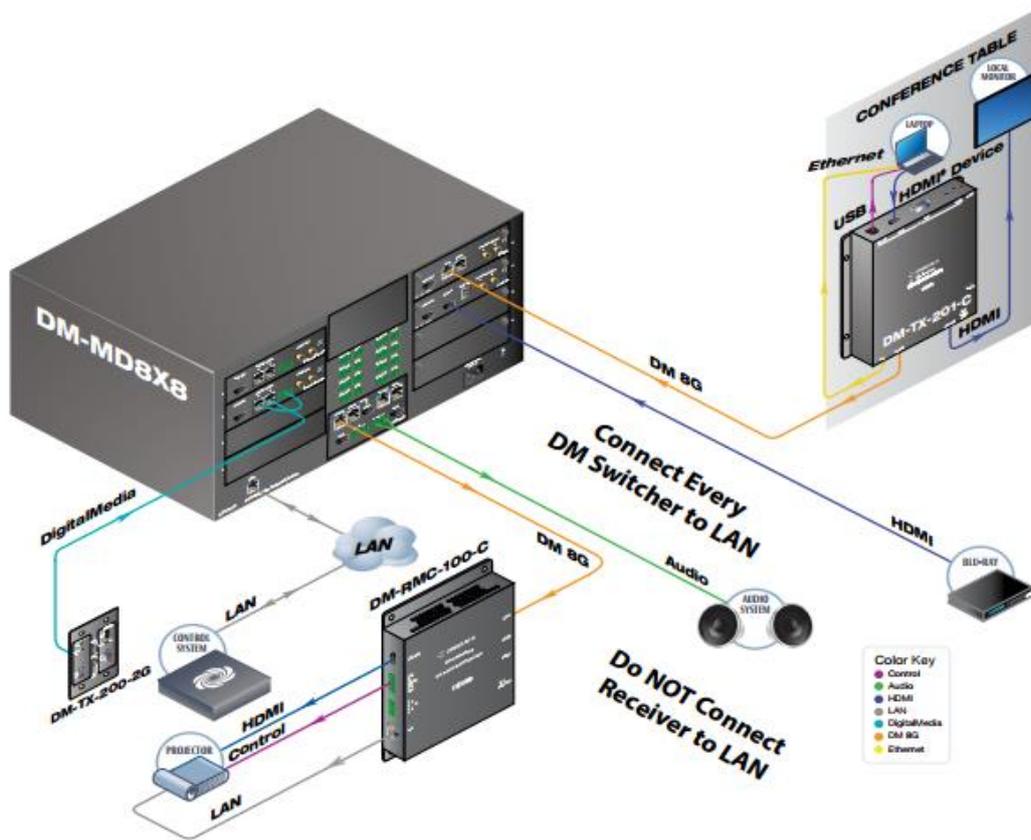


Figura 15. Diagramma di un'installazione multifunzione (www.crestron.com)

La scelta dell'una o dell'altra tecnologia può dipendere da molti fattori, di cui quello economico è spesso quello non determinante dato che si collocano su livelli di qualità paragonabili.

Un fattore importante è la "consuetudine", ovvero spesso certi mercati (per esempio cantieri navali) hanno abbracciato una famiglia di prodotti e sono restii a cambiarla data la confidenza tecnica raggiunta e gli accordi commerciali che si sono via via negli anni instaurati.

Un altro fattore è la presenza di "centri assistenza" e personale formato in loco per la realizzazione di interventi rapidi e modifiche in caso di necessità.

Ultimo fattore è il design hardware delle interfacce di controllo utente che spesso differiscono in maniera sostanziale.



Figura 16. Panoramica touch panel Crestron



Figura 17. Panoramica touch panel Control4

2.2 Le tecnologie open source

Accanto alle tecnologie professionali si affiancano approcci basati su componenti open-source che riescono a garantire una buona affidabilità con una considerevole diminuzione dei costi.

Il rovescio della medaglia è la questione di manutenibilità del progetto nel tempo e gestione delle assistenze; questi tipi di soluzioni sono spesso custom-made e il mercato professionale (che spesso punta ai grandi marchi anche per una questione di prestigio dell'installazione stessa) spesso le rifiuta, ma sono ottimali per progetti hobbistici e didattici.

2.2.1 Arduino

Arduino è una scheda elettronica di dimensioni ridotte con un microcontrollore, utile per creare rapidamente prototipi o per scopi hobbistici, didattici e consumer.

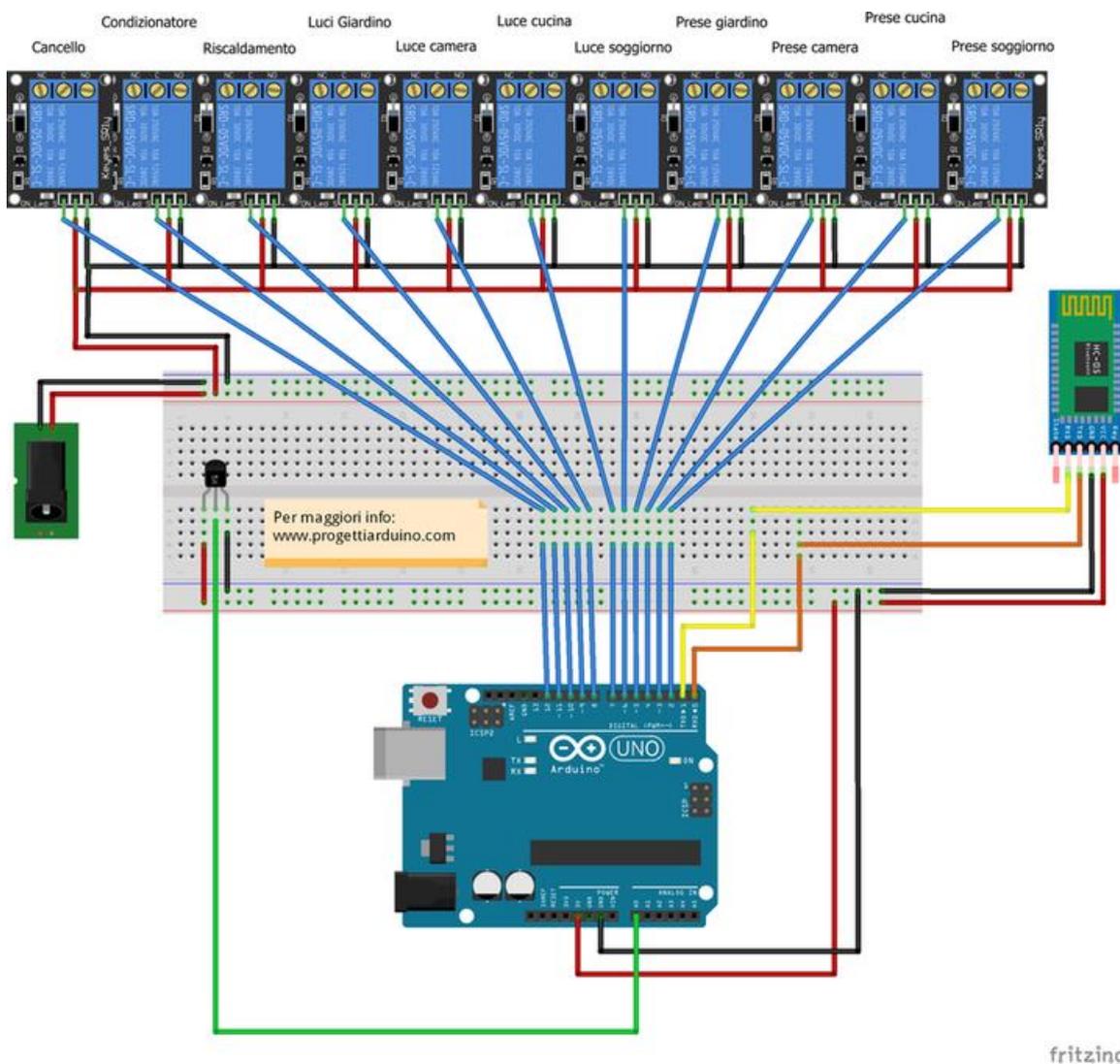


Figura 18. Microcontrollore Arduino (fonte www.progettiarduino.com)

Tutto il software a corredo è libero (si programma in *Wiring*, linguaggio derivato dal C e dal C++), e gli schemi circuitali sono distribuiti come hardware libero.

Non è volontà di questo elaborato entrare nel dettaglio della soluzione e si rimanda ai riferimenti [7] per eventuali approfondimenti.

C'ho che Arduino mette a disposizione, oltre che il controllore con la sua logica programmata, è anche la componente hardware specifica.

Risulta comunque possibile impiegare una qualunque macchina dedicata di qualsiasi sistema operativo (Windows, Linux, ecc.) per creare il proprio controllore intelligente (programmandola ad esempio con linguaggi free tipo Python, C++) limitato al controllo di dispositivi via LAN o seriale (oppure si può estendere l'installazione con l'utilizzo di extender seriali ed infrarosso pilotati via LAN tramite apposite API).

2.3 Le tecnologie "plug and play"

In questo paragrafo saranno elencate alcune delle soluzioni che spesso sono etichettate come "plug and play" per dar enfasi al fatto che non necessitano di alcuna programmazione alle spalle per poter essere implementate.

Naturalmente questa mancanza di personalizzazione e di controllo ne limitano l'utilizzo in ambito domestico personale e sono state inserite nell'elaborato solo per completezza e per mostrare su quanti livelli diversi, la domotica, si possa presentare e percepire.

Visti i grandi brand alle spalle sono tecnologie importanti nella loro funzione di "traino ed attrattiva" al tema domotico di un pubblico più vasto. È personale opinione che non vadano ad intaccare il mercato delle tecnologie professionali, ma anzi vi facciano migrare numerose persone che, una volta capita l'utilità della tecnologia, decidano di passare a soluzioni di più alto livello.

2.3.1 Apple Home Kit

Apple recentemente si è affacciata al mercato domotico (visto forse il forte mercato in crescita) presentando la app *Casa* e la tecnologia Home Kit.

La soluzione permette di gestire diverse tipologie di apparati di uso consumer tramite un'unica App dedicata. È possibile creare un minimo di logica per far "dialogare" le tecnologie arrivando a creare semplici scenari personalizzati.

Accessori HomeKit. Una famiglia sempre più grande. E intelligente.

Aziende leader in tutto il mondo offrono accessori compatibili con HomeKit e con i tuoi dispositivi Apple, e il loro numero continua a crescere. Dai un'occhiata a questa pagina di tanto in tanto, per vedere se ci sono novità.

Cerca questo simbolo nei negozi e online.



Luci



Interruttori



Prese elettriche



Termostati



Finestre



Ventilatori



Condizionatori



Sensori



Sicurezza



Serrature



Videocamere



Campanelli



Porte del garage



Bridge

Figura 19. Apple HomeKit: lista accessori compatibili

L'applicazione permette alcune piccole personalizzazioni come ad esempio la scelta dell'immagine di sfondo dell'interfaccia utente.

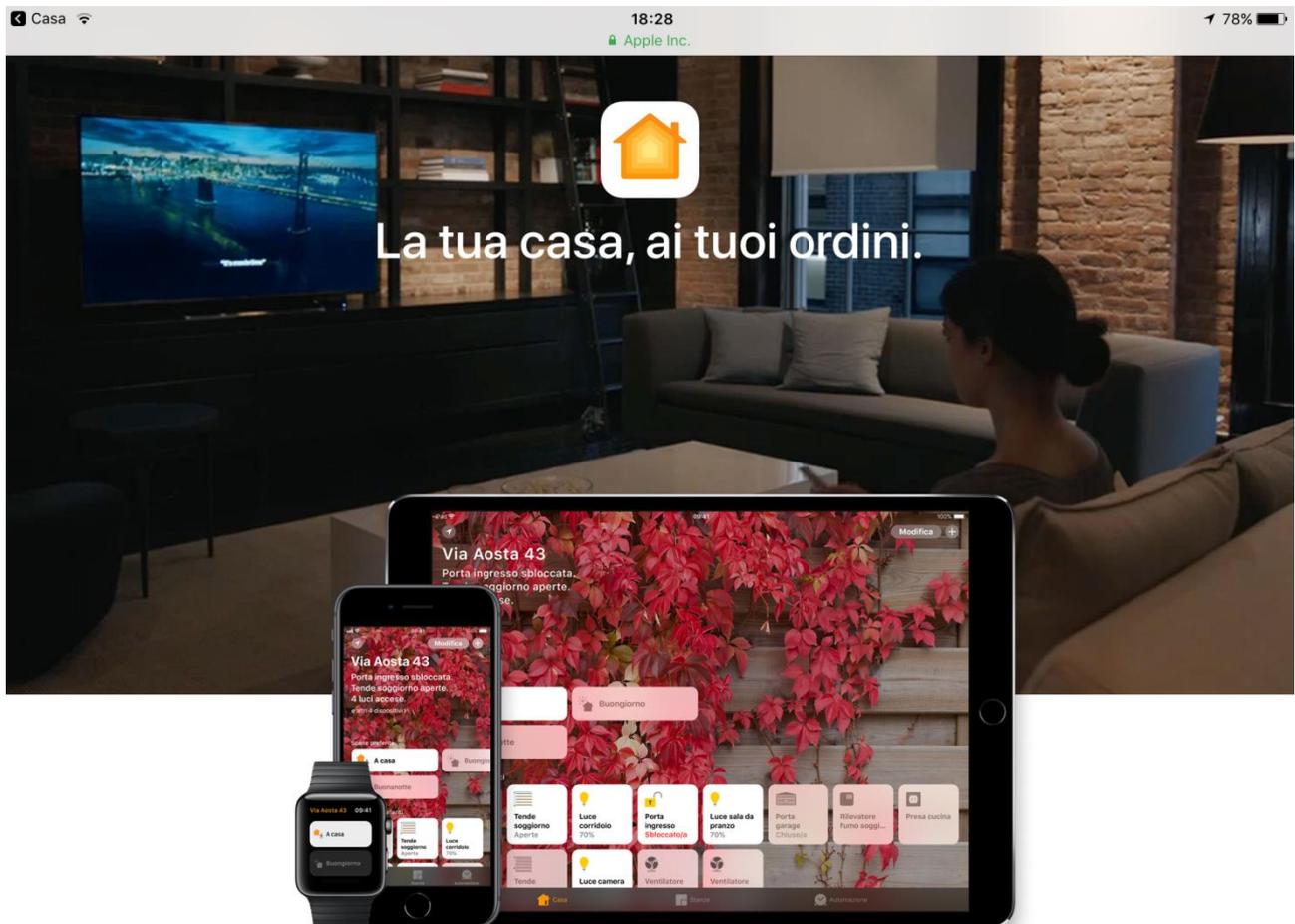


Figura 20. Apple HomeKit: interfaccia di controllo

Particolare importanza viene data al controllo vocale tramite Siri, l'assistente vocale di casa Apple.

2.3.2 Google Home

Anche Google si sta affacciando al mercato domotico nell'insieme di tecnologie e soluzioni che stanno sotto il cappello di Google Home. Nata inizialmente per la gestione di contenuti audio, recentemente è stata anche destinata alla home automation tramite input vocale.



Figura 21. Dispositivo Google Home

Anche in questa tecnologia, dunque, viene data grande rilevanza al controllo vocale (motivo per il quale ne verrà dedicata una sezione specifica) gestita dall'assistente vocale Google Assistant (per ora non disponibile in italiano).

2.3.3 IFTTT

Per concludere la panoramica delle soluzioni che in qualche modo abbracciano il concetto di domotica IFTTT (*If This Then That*) ha una sua importanza specifica.



Figura 22. IFTTT

IFTTT (se questo allora quello) fa parte di quelle tecnologie definite *End User Development* ovvero quelle soluzioni che non necessitano di alcuna conoscenza sistemistica o di programmazione per poter essere messe in funzione.

IFTTT è un ambiente (web o mobile) che permette di creare delle "ricette" in funzioni di eventi e trigger. Il *questo* (trigger) può essere di varia natura (es. Check-in in Foursquare, input da geolocalizzazione) e il *quello* (l'azione) può essere l'invio di un messaggio di testo, la gestione di luci (es Philips Hue) o termostati (es. Nest) e molti altri generi di attuazioni.

Sono possibili configurazioni di logiche e condizioni combinate (AND e OR) per raggiungere un minimo di personalizzazione di funzionalità.

Per curiosità ed approfondimenti si rimanda a [8].

Come Google Home e Apple Home Kit queste tipologie di approcci hanno piucchealtro una valenza propedeutica e di avvicinamento alla domotica che non di impianto domotico “vero” e professionale.

Fanno percepire inoltre il concetto di *internet of things*, ovvero “l’estensione” del web al mondo reale e viceversa.

3 Il Controllo Vocale

La centralità del controllo vocale data dai grandi brand descritti in precedenza ne sottolineano l’importanza e l’attualità anche in ambito domotico.

In questa sezione verranno introdotte le motivazioni che spingono a considerare il controllo vocale all’interno dei sistemi domotici e delle criticità che questi soffrono e che ne limitano a tutt’oggi l’utilizzo.

Saranno presentati prima alcuni concetti introduttivi al fine di una migliore collocazione tematica.

3.1 Multimodalità

Nell’interazione uomo macchina una **modalità** è un modo di interagire con un sistema che sfrutti uno specifico senso umano [9].

Quindi un’interfaccia è multimodale se sfrutta più sensi per interagire (vista, udito, gesti, espressioni, ecc).

Il **dialogo fra persone** (comunicazione faccia a faccia) è in genere multimodale [2]; nella multimodalità gli elementi non verbali (i segni paralinguistici quali espressioni del volto, gesti, posizioni del corpo, ecc.) sono importantissimi: l’uso frequente dei riferimenti deittici (espressioni come “questo” o “quello” accompagnato o esclusivamente referenziato con un gesto) ne sono un esempio lampante.

Il linguaggio verbale (le parole che usiamo) rappresenta solo una minima parte della nostra capacità di comunicazione. La componente non verbale è spesso la componente principale.

Noi esseri umani non ce ne accorgiamo perché per noi è naturale fruire l’informazione da più canali, farne sintesi e operazioni inferenziali e soprattutto “risolvere” le ambiguità in base al *contesto*.

Tutte queste operazioni non sono possibili (o lo sono solo in parte) da una sistema di riconoscimento vocale che si può basare solo sull’input vocale: ecco perché tale approccio risulta spesso difficoltoso e non attuabile se non sotto particolari condizioni.

Per capire le motivazioni che spingono allo studio del controllo vocale (come di altri modi oltre a quello visivo) va fatto un passo indietro per comprendere i vantaggi della multimodalità.

Per fare questo partiamo dall’uomo e dai suoi processi cognitivi.

L’uomo ha tre tipi di memoria [9]:

- Sensoriale
- A breve termine (o di lavoro)
- A lungo termine

Queste tre tipologie trattengono l'informazione per una durata di tempo variabile (ordine del decimo di secondo per la sensoriale, del secondo per quella di lavoro, ordine di giorni, mesi o anni per quella a lungo termine).

L'informazione in memoria a lungo termine è codificata in maniera diversa a seconda del canale sensoriale attraverso cui giunge al destinatario. La letteratura parla di **superiorità dell'immagine** (teoria doppia codifica) e ne porta prove inconfutabili a riguardo.

La multimodalità risulta utile nel suddividere l'informazione su più sensi *per migliorare l'efficienza della memoria di lavoro* (gli esseri umani sono in grado di elaborare modalità in modo parzialmente indipendente e quindi, presentando informazioni con molteplici modalità, aumenta la capacità della memoria di lavoro [1]), per aumentare l'usabilità e per questioni di accessibilità.

3.2 Voice User Interface (VUI)

Le interfacce vocali soffrono, come risaputo, del problema della *transitorietà dell'informazione* (altra caratteristica della lingua parlata [2]) e del problema sulla memoria di lavoro a breve termine precedentemente indicato.

Studi riguardo quest'ultima asseriscono che mediamente la memoria a breve termine dell'indivio comune si aggira intorno ai 7 elementi $+2 - 2$ (da 5 a 9). Questo buffer è un limite da tenere in considerazione quando si progetta un qualunque sistema interattivo.

Questi due elementi rendono la progettualità dei sistemi di controllo vocale estremamente complessa; i concetti di usabilità e di "buon design" analizzati in precedenza devono essere implementati con maggior rigore (oltre che rivisitati in funzione delle peculiarità del canale) se vogliamo che la soluzione sia fruibile dall'utilizzatore.

L'architettura dell'informazione avrà un ruolo estremamente preponderante dato che il successo di un'interfaccia vocale dipenderà sia dal numero delle opzioni proposte ma anche dalla logicità della loro organizzazione.

Risulterà inoltre essenziale permettere all'utente di selezionare l'opzione desiderata in un qualunque momento, interrompendo la lista se necessario e di richiedere la ripetizione di una o più opzioni senza doverle riascoltarle tutte (controllo e libertà, flessibilità: Nielsen).

Tutte questi approcci e soluzioni tipiche di chi si rivolge allo sviluppo di un'interfaccia vocale sono racchiuse nel termine **VUI (Voice User Interface) Design** [3].

Una metodologia proficua è quella di restringere, ad ogni passo di interazione, le possibilità di scelta grazie all'uso di *vincoli*, fino ad arrivare, ove possibile, al semplice "si" e "no".

Ad esempio si prendano i due output di un sistema vocale:

- "Per favore conferma che il numero di interesse è 123-456"
- "Il numero di interesse è 123-456? Per favore rispondi con un 'si' o con un 'no'".

Le risposte umane al primo approccio possono essere molteplici:

- "Sì, giusto il numero è 123-456"
- "Sì"
- "Ok"
- Ecc.

L'unica risposta concessa al secondo approccio è "sì" o "no" (in mancanza di questi input il sistema potrà richiedere la ripetizione del comando).

Un'altra tecnica è quella di includere nella domanda le risposte possibili, invitando l'utente (tramite questo suggerimento implicito) a dare la risposta corretta, come ad esempio:

Domanda: "cosa vuole fare accendere la luce, attivare lo scenario o accendere lo stereo?"

Risposta umana: "attivare lo scenario".

L'esperienza ha mostrato come un'attenta progettazione VUI sia spesso più proficua in termini di usabilità (efficacia, efficienza e soddisfazione) che il semplice miglioramento delle performance (riduzione del Word Error Rate) ottenuto tramite l'ottimizzazione degli algoritmi di riconoscimento vocale.

Nonostante tutti gli sforzi c'è una grossa limitazione organica delle interfacce vocali, ovvero quella che viene definito il **"problema di scopribilità"** (Nielsen) [10];

Un po' come le interfacce prompt a linea di comando dei computer di qualche decennio fa, le funzionalità non sono presentate in modo esplicito come avviene invece con un semplice menu grafico visivo, ma il sistema va "interrogato" a priori (in analogia con il comando "dir" o "ls" rispettivamente su Windows e linux).

3.3 Violazione dei principi di usabilità e di design

Come visto nel paragrafo precedente il controllo vocale soffre di molte problematiche che un progettista non può ignorare:

- Il problema della scopribilità -> cosa può fare il sistema? A quali comandi può rispondere?
- Problema della visibilità dello stato del sistema (violazione della prima euristica di Nielsen) che si lega indissolubilmente al ->
- Problema del sovraccarico della memoria a breve termine

Si incontrano inoltre altre violazioni [4]:

- Violazione dell'euristica "Favorire il riconoscimento piuttosto che il ricordo"

"Supporting recognition over recall is also severely constrained in a voice-only interface, since even reciting a list of options requires users to store the options in working memory while they make a selection."

- Violazione dell'euristica "prevenzione dell'errore": l'errore è innato in un sistema di riconoscimento vocale. Infatti nonostante tutti gli sforzi fatti nella ricerca e sviluppo in tutte le tecnologie (Reti Neurali Ricorrenti, Catene Markov Nascoste, ecc) il Word Error Rate (WER) non è mai diverso da zero.

"One of the classic usability heuristics is error prevention: the notion that rather than just help users recover from errors, systems should prevent errors from occurring in the first place. As speech recognition has improved in recent years, errors in natural-language understanding have dropped significantly. Hopefully this trend will continue for all voice interaction systems."

Anche se si parla di pochi punti percentuali (intorno 5%) siamo enormemente al di sopra delle percentuali di errori sul comando grafico (si sbaglia un click sul bottone meno di una volta 1 su mille).

3.4 Una tappa inevitabile?

Sembrano veramente tanti i *contro* all'impiego della voce come metodo di input verso un qualsiasi sistema informatico: violazione di principi base di usabilità e di design interattivo, difficoltà organica dell'interpretazione delle parole, maggior complessità nella definizione dell'interfaccia e dell'architettura informativa associata.

L'unico *pro* finora incontrato consiste nella diminuzione del carico cognitivo grazie alla separazione dell'informazione su un canale diverso da quello visivo (sempre che anche il feedback sul comando effettuato interessi il canale uditivo, altrimenti è solo un miglioramento parziale).

In realtà il confronto non è finito, e per pareggiare i piatti della bilancia bisogna spostarci ad un livello *più profondo* che gli aspetti "tecnici ed informatici" non contemplano ma quelli umanistici sì.

La comunicazione di tipo orale è la forma più innata e comune che ognuno di noi sviluppa in maniera naturale senza bisogno di insegnamenti, semplicemente "vivendo" ed "interagendo" in un determinato ambiente sociale e culturale.

Alla nascita dei computer sembrava naturale che il metodo di comunicativo con essi dovesse essere la parola e non la scrittura, essendo quest'ultima da sempre specifica per la memorizzazione dell'informazione e non tanto allo scambio dell'informazione, ruolo privilegiato, appunto, dall'oralità.

Purtroppo le difficoltà tecniche del riconoscimento del parlato incontrate fin dagli esordi dei computer hanno portato ad una "virata" verso la scrittura e successivamente alle immagini come tecniche privilegiate di interazione uomo-macchina.

Questo non deve però fare gettare la spugna perchè prima o poi, quando le tecniche e gli approcci (vedi VUI Design) saranno maturi sarà inevitabile la predilezione verso il controllo vocale.

Alcuni principi generali cui tante discipline (letteratura, marketing, ecc.) ed ambiti tecnologici sono andati incontro nel corso degli anni sono racchiuse nelle "lezioni americane" di **Italo Calvino**.

Italo Calvino, fu invitato ad Harvard per un ciclo di conferenze da tenere nell'anno accademico 1985-1986. Purtroppo queste conferenze non furono mai tenute a causa della morte stessa dello scrittore nel 1985.

In queste "lezioni americane" egli sviluppò sei idee che ribattezzò "6 proposte per il nuovo millennio" che riteneva essere alla base della letteratura e della comunicazione per gli anni a venire.

1. **Leggerezza:** Tutto ciò che è leggero è vincente e per leggero si intende sia leggero in senso fisico che in senso pratico.
L'iPod ha vinto lo stereo, i tablet hanno vinto i computer.
Le piattaforme e i servizi web-based vincono sui software standalone.
Gli e-book vincono i libri di carta.
L'archiviazione su cloud è vincente rispetto agli hard disk.
Lo streaming audio-video batte il supporto fisico.
La voce vincerà sui comandi visivi.
2. **Rapidità:** Siamo nell'epoca della rapidità. La soluzione più veloce, che non implica troppi preamboli, vince.
3. **Esattezza:** intesa come un progetto ben definito e studiato, che non si fermi di fronte alle difficoltà ma che cerchi di aggirarle inventando nuove soluzioni ed approcci.
4. **Visibilità:** come capacità di essere messo in vista, di essere promosso, di essere eletto come migliore. Basti pensare all'incremento dell'utilizzo dei messaggi vocali (whatsapp) a discapito di quelli testuali.

5. **Molteplicità:** L'uso molteplice di più canali di comunicazione interconnessi fra loro da uno schema preciso di relazione e interrelazione deve creare un'armonia ed unità alla comunicazione.
6. **Coerenza:** esseri coerenti con tutti gli altri principi e tenere a mente le motivazioni e le condizioni in cui si opera.

Risulta chiaro come il controllo vocale abbracci tutti questi principi ed è quindi inevitabilmente una strada da percorrere e sulla quale investire tempo e denaro.

3.5 Comando vocale: un ambito di studio aperto

Se certi aspetti sembrano proclamare il controllo vocale come il futuro per la gestione delle interfacce, altri lo limitano ad ambiti circoscritti e limitati.

Ad oggi l'uso esclusivo di interfacce vocali è sconsigliato in contesti critici dove l'immediatezza dell'informazione o la situazione emotiva dell'utilizzatore diventano predominanti nell'atto interattivo (ad esempio stress dell'utente dovuto alla perdita della carta di credito).

Si preferisce quindi integrare il parlato con altro modalità di fruizione dell'informazione (ad esempio visiva).

Il peso della comunicazione vocale all'interno della comunicazione multimodale insieme allo studio di convenzioni comunicative (VUI e computer talk) costituiscono temi cardini della ricerca attuale sull'interazione uomo-macchina.

Parte II: Domotica e controllo vocale

Come descritto nel paragrafo introduttivo gli input attuativi per un sistema domotico possono essere molteplici dalla semplice chiusura di un contatto elettrico, al click su un'interfaccia grafica, la variazione di stato di un sensore, ecc.

L'escursus fatto nella precedente sezione relativa al controllo vocale è da ritenersi come introduttivo e propedeutico per questa ultima parte dell'elaborato dedicata ad un'implementazione pratica ovvero il controllo vocale di una tecnologia domotica, ovvero l'integrazione del toolkit CMUSphinx [5] con il sistema di controllo luci Philips HUE [6].

4 CMUSphinx descrizione architettura

Sphinx è un sistema di riconoscimento vocale automatico con le seguenti caratteristiche:

- Indipendenza dal parlante
- capacità di riconoscimento in parlato continuo
- gestione vasto vocabolario

Sphinx è basato su Modello Marcoviani Nascosti (HMM) come il popolare Dragon Naturally Speaking e come questo non ha bisogno della connettività di rete (Internet) per poter operare dato che lavora esclusivamente in locale (evitando così qualsiasi problema di violazione di privacy potenzialmente presente nelle soluzioni Apple Siri e Google Home) .

Attualmente le percentuali di riconoscimento per la lingua inglese sono state stimate con un WER intorno al 15%.

Il processo di riconoscimento vocale si suddivide nei seguenti passi:

1. Acquisizione del parlato e sua trasformazione da oscillogramma a *spettrogramma* passando quindi da una rappresentazione su base temporale e base frequenziale.
2. Lo spettrogramma viene digitalizzato e quantizzato ottenendo in uscita una sequenza di *feature vector*.
3. Con la fase di *segmentazione* si passa dalla sequenza continua del parlato all'individuazione sullo spettrogramma dei singoli foni costituenti le parole e la discriminazione di ciò che è parlato o meno (si escludono rumore di fondo e momenti di silenzio fra la singole parole)
4. Fase di *decodifica* delle parole.

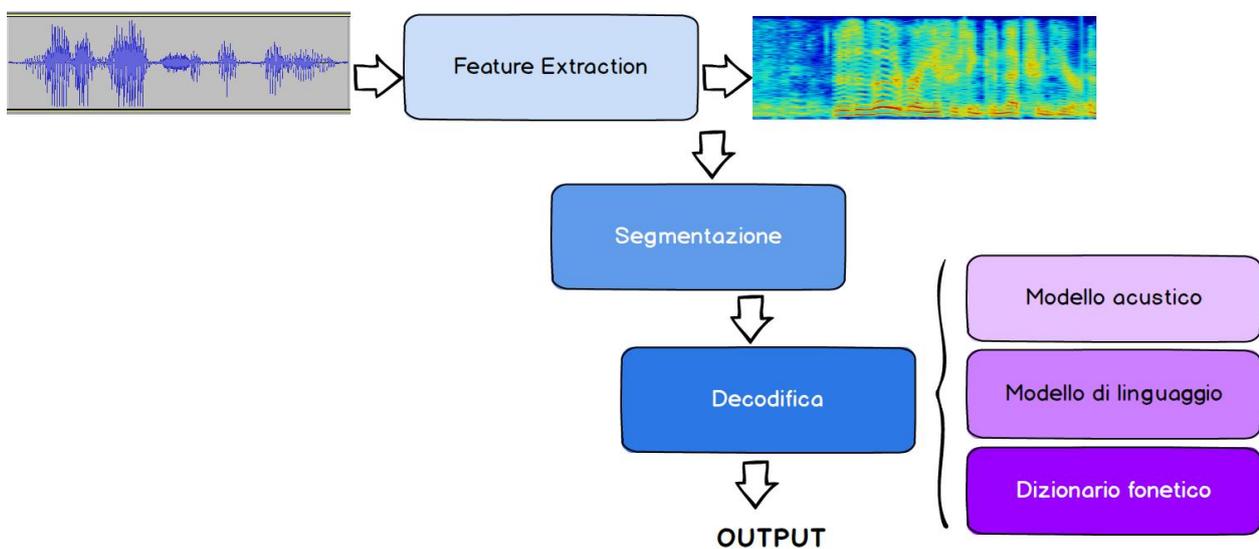


Figura 23. Il processo di riconoscimento vocale

Come descritto in [3] la decodifica del parlato si può descrivere nel seguente modo:

Un modello statistico del linguaggio basato sui *trigrammi* come una macchina a stati finiti (Catena Markoviana) costituita da parole; le parole a loro volta sono rappresentate da una macchina a stati finiti costituita da fonemi; i fonemi possono essere rappresentati ancora come un'ulteriore macchina a stati finiti costituita da stati HMM.

In conclusione la decodifica avviene attraverso una grande macchina a stati finiti dove le transizioni sono descritte da probabilità.

Nel toolkit Sphinx tutto questo si basa su tre componenti:

- 1 Modello Acustico (Acoustic Model):** mappa le relazioni fra suoni e foni. Ogni fono è suddiviso in 3 stati (iniziale, centrale e finale).

Nel modello non si considera un singolo fono ma la sequenza di tre foni in un contesto (il precedente ed il successivo a quello attuale) il tutto per gestire la *coarticolazione* vocale [3].

Il componente è specifico per ogni lingua (Italiano, Inglese, ecc.) e per tipologia di canale (telefonico, broadcast, ecc.).

Può essere adattato a condizione acustiche ed accenti particolari.

Nello specifico ogni modello acustico contiene i Modelli Markoviani Nascosti associati ai singoli foni.

Il “code-mixing” (ovvero l’utilizzo di diverse lingue in una frase) può essere gestito a questo livello integrando il modulo acustico di due lingue diverse

<https://cmusphinx.github.io/wiki/tutorialadapt>.

Il motivo per cui i sistemi di riconoscimento vocale danno risultati peggiori se sottoposti ad un parlante non madrelingua può essere associato ancora a questo livello: l’articolazione vocale dei loro foni non sarà “perfetta” come per parlante L1 e l’interpretazione del modello acustico ne risentirà.

2 Dizionario Fonetico (Phonetic Dictionary): contiene il mapping fra parole e la loro pronuncia composta dai singoli fonemi: il cosiddetto “phonetic spelling”.

Lo standard seguito è quello [ARPAbet](#).

Ad esempio un estratto dal pronunciation-dictionary inglese si presenta come segue:

abboud AH B UW D
abbreviate AH B R IY V IY EY T

Il suo ruolo è quello di aggiungere al modello parole altrimenti sconosciute.

Il Dizionario insieme al Modello di Linguaggio è basilare per estendere il numero di parole riconoscibili dal sistema ed è utile ad esempio per particolari applicazioni (es. Controllo traffico marittimo dove è richiesta una terminologia specifica).

La parola aggiunta deve essere contemplata anche nel Modello di Linguaggio altrimenti non verrà mai riconosciuta.

Ad ogni parola possono essere aggiunte varianti di pronuncia (causate ad esempio dalla coarticolazione o da piccole variazioni dovute alle differenze dei tratti vocali dei singoli parlanti); basta indicare le possibili varianti tra parentesi tonde ed ordine progressivo come ad esempio:

abbreviated AH B R IY V IY EY T AH D
abbreviated(2) AH B R IY V IY EY T IH D

3 Modello di Linguaggio (Language Model): Definisce le sequenze di tutte le possibili parole che possono essere interpretate.

Il suo ruolo è sia quello di ampliare le possibili sequenze di parole sia quello di restringere le possibilità alle sole contemplate nel modello stesso.

Con un approccio a macchina a stati finiti (Catena di Markov) “cattura” la sequenza di parole usando il modello a “tri-gramma” (la probabilità di una parola dipende dalle due precedenti) lavorando su base statistica.

Esistono infatti grammatiche statistiche che associano la probabilità di accadimento di ogni sequenza di tre parole.

Si parte dalle prime 3 parole e le loro alternative possibili ed incrociandole con queste grammatiche statistiche si estrae la sequenza più probabile.

Si procede spostandosi di una parola più a destra e si ripete il procedimento.

Alla fine si otterrà la sequenza di parole (composta da spezzoni di 3) più probabile.

Questa sarà l’output del sistema di riconoscimento.

Il Modello di Linguaggio può essere particolarmente utile per “catturare” la sequenza di parole usate in un particolare dominio (es navale, aereoportuale, medico) e per disambiguare parole identiche dal punto di vista fonico (es “disc” e “disk”).

Un estratto dal language model inglese si presenta come segue (i numeri che precedono e seguono la sequenza sono pesi associati alla sequenza stessa):

\1-grams:

-5.6332 abscess -0.2435

-6.3384 abscond -0.1849

-6.3074 absconded -0.2692

-6.3352 absconding -0.2496

-4.6783 absence -0.5733

\2-grams:

-3.6058 autopsy receiving 0.0000

-2.7527 autopsy records 0.0000

-1.1515 autopsy report -0.0912

-3.6213 autopsy report's 0.0000

-1.6965 autopsy reports -0.0464

\3-grams:

-1.2699 having waited twelve

-0.6122 having was screaming

-1.2368 having washington spend

-2.0820 having watched fung

-1.2117 having watched him

Il Language Model è il blocco “dell’intelligenza semantica e sintattica”.

Solo in questo blocco possiamo stabilire le relazioni tra sequenze di parole, ossia è qui che possiamo definire sia “*gruppi semantici*”, parole che ne portano altre dietro perché connesse a livello di significato, sia i “*gruppi sintattici*”, ovvero la possibilità di stabilire regole grammaticali per l’output del processo di riconoscimento.

4.1 Analisi del modello attuale per l'italiano

Il modello di default implementato per l'italiano deriva da [VoxForge](#).

Per la lingua italiana i fonemi considerati nel modello del canale, in formato ARPAbet, sono 38.

EE	a1	f	m	r	v
JJ	b	g	n	s	w
LL	d	i	nf	t	z
OO	dZZ	i1	ng	tSS	
SIL	dz	j	o	ts	
SS	e	k	o1	u	
a	e1	l	p	u1	

Il numero di parole singole presenti nel Language Model (1-grams) sono circa 3700.

Nel dettaglio:

ngram 1=3741

ngram 2=11194

ngram 3=13364

Questo numero risulta molto inferiore al "vocabolario di base" italiano ovvero "l'insieme dei vocaboli necessari per capire e farsi capire nelle situazioni comunicative più frequenti" che comprende circa 6700 lessemi [fonte GRADIT].

A maggior ragione risulta irrisorio se confrontato col "vocabolario corrente" ovvero circa 50000 lemmi (vocabolario di base sommato a quello corrente) che rappresenta "l'insieme dei termini generalmente noti a chiunque abbia un livello medio superiore di istruzione"[De Mauro 1999, XX]

Il vocabolario esteso italiano, che comprende anche i tecnicismi, regionalismi e dialettismi, si aggira intorno ai 250000 lemmi.

Analizzando i file di dizionario fonetico e del language model è evidente come numerose parole che potrebbero essere utili in campo domotico non siano contemplate.

Risulta inevitabile ai nostri scopi la creazione di un nuovo language model per l'italiano cercando di colmare questa "carezza lessicale" del modello attuale disponibile. Non verranno prese in considerazione variazioni ed integrazioni al Channel Model in quanto ritenuto "sufficientemente buono" dopo una prima fase preliminare di test.

Per gli step implementativi si rimanda a [5].

4.2 Descrizione Philips HUE

Come accennato a riguardo della figura del system integrator, le informazioni sul controllo del dispositivo vengono fornite dai produttori sotto forma di API.

La via di accesso della comunicazione Philips è il bridge HUE che grazie all'interfaccia [RESTful](#) si presenta come un servizio web gestibile attraverso API dedicata.

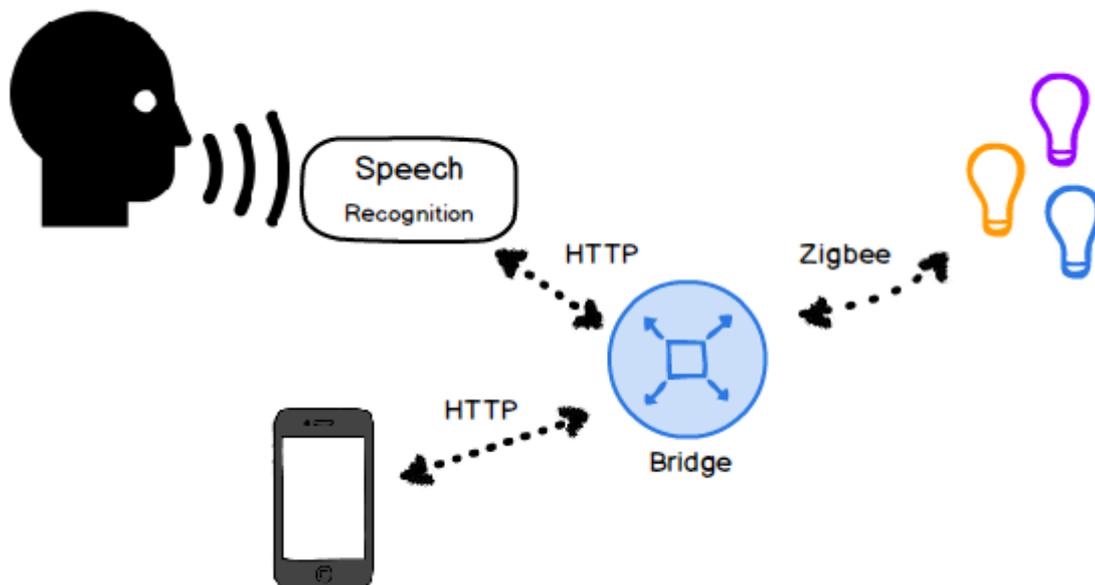


Figura 24. Diagramma Controllo vocale e tecnologia Philips Hue

La descrizione completa dell'API è consultabile al link <https://www.developers.meethue.com/documentation/getting-started> previa registrazione.

Un breve estratto dei comandi attuabili è il seguente:

- Spegnimento luce 1:

Address	<code>http://<bridge ip address>/api/[KEY]/lights/1/state</code>
Body	<code>{"on":false}</code>

Address	<code>http://<bridge ip address>/api/[KEY]/lights/1/state</code>
Method	PUT

- Accensione luce 1 e settaggio colore specifico:

Address	<code>http://<bridge ip address>/api/[KEY]/lights/1/state</code>
Body	<code>{"on":true, "sat":254, "bri":254,"hue":10000}</code>
Method	PUT

4.3 Creazione Language Model ad hoc

I comandi a cui vogliamo che il sistema risponda sono:

- accendi/spengi/spegni luce/luci [in] camera/bagno es. “accendi luce in bagno”
 - luce [in] camera/bagno accesa/spenta es. “luce bagno spenta”
- dove [] indica parametro opzionale.

Il Dizionario Fonetico, per la parte specifica di interesse, si presenta così:

accendi a ttS E1 n d i

bagno b a1 JJ o

camera k a m e r a

in i1 n

luce l u1 tS e

luci l u1 tS i

spegni s p e1 JJ i

spengi s p e1 n dZ i

Il Language Model sarà costituito dai seguenti n-grammi che includono le combinazioni delle varie sequenze di parole nel comando vocale:

- ngram 1=12
- ngram 2=26
- ngram 3=52

4.4 Verifica funzionamento

Al blocco di riconoscimento del parlato è stato messo in cascata un “blocco di comprensione” che attraverso un *parsing robusto* [3] interpreta le *parole chiave* di luogo (camera/bagno) e azione (accendi/spengi) e genera le chiamate RESTful al bridge HUE.

Come feedback ai comandi inviati segue una conferma vocale così da mantenere tutta la gestione su un vero e proprio modo indipendente ed autonomo (canale completamente separato da quello visivo).

I test, effettuati da 3 persone in ambiente domestico in modalità “riconoscimento continuo del parlato da microfono”, hanno riportato una percentuale di completamento del compito superiore al 90%.

Le alte percentuali di completamento del compito sono da attribuire sia al fatto che il sistema interpreta solo determinate *parole chiave* specifiche al contesto sia alla predisposizione dell’utente ad interagire col controllo vocale utilizzando una versione semplificata del parlato naturale, che potremmo definire “*computer talk*”, caratterizzata da una dizione più chiara e lenta delle parole rispetto alla normale pronuncia.

L’ esempio mostra come le qualità di un sistema di riconoscimento vocale debbano essere sempre contestualizzate alle funzionalità reali del prodotto e degli utilizzatori finali.

5 Commenti conclusivi

L’elaborato è partito dando nozioni di domotica e presentando un overview dei principali ambiti di impiego per poi passare a concetti di interazione uomo-macchina e principi universali di design interattivo.

Tra i metodi di interazione è stato preso in considerazione il controllo vocale del quale ne sono stati descritti pregi e difetti. Alla fine il giro si è chiuso mostrando un’applicazione domotica gestita da un sistema di riconoscimento vocale open source.

6 Riferimenti bibliografici

- [1] Don Norman, *La caffettiera del masochista*
- [2] Miriam Voghera, *Lingua parlata*, Treccani
- [3] Roberto Pieraccini, *The voice in the machine*
- [4] Jacob Nielsen, <https://www.nngroup.com/articles/voice-interaction-ux/>
- [5] <https://cmusphinx.github.io/>
- [6] <https://www.developers.meethue.com/philips-hue-api>
- [7] [https://it.wikipedia.org/wiki/Arduino_\(hardware\)](https://it.wikipedia.org/wiki/Arduino_(hardware))
- [8] <https://ifttt.com>
- [9] Gamberini, Chittaro, Paternò, *Human-Computer Interaction - I fondamenti dell'interazione tra persone e tecnologie*
- [10] Mirko Tamosanis, *La lingua e l'intelligenza artificiale*