

# SiVoDiM

---



## Verbale 2016-03-09

<b>Versione</b>	1.0.0
<b>Redattori</b>	Francesco Bizzaro, Gino Zaidan
<b>Verificatori</b>	Riccardo Rizzo
<b>Responsabili</b>	Alberto Andriolo, Enrico Chiara
<b>Uso</b>	Esterno
<b>Lista di distribuzione</b>	<i>Stark Labs</i> Prof. Vardanega Tullio, Dr. Cardin Riccardo

Verbale del primo incontro in data 09-03-2016 con Giulio Paci.

## Registro delle modifiche

Attività	Autori	Data	Versione
Approvazione documento	Enrico Chiara, Alberto Andriolo	10-03-2016	1.0.0
Verifica documento	Enrico Chiara, Alberto Andriolo	10-03-2016	0.2.0
Verifica documento	Riccardo Rizzo	10-03-2016	0.1.0
Stesura del documento	Francesco Bizzaro, Gino Zaidan	09-03-2016	0.0.2
Creazione struttura documento	Francesco Bizzaro	09-03-2016	0.0.1

## Indice

<b>1</b>	<b>Informazioni Generali</b>	<b>1</b>
<b>2</b>	<b>Questioni emerse</b>	<b>1</b>
2.1	Quale obiettivo si vuole raggiungere con l'applicazione? . . . . .	1
2.2	Per quale motivo si è scelto il mondo mobile? . . . . .	1
2.3	Si richiede che l'applicazione elabori una frase in tempo reale, senza usare audio creato precedentemente? . . . . .	1
2.4	Supposto di utilizzare il motore di sintesi in un'applicazione di messaggistica, quali sarebbero i vantaggi offerti rispetto all'invio di note vocali? . . . . .	2
2.5	Quali sono le principali problematiche riguardanti la connessione al servizio remoto? . . . . .	2
2.6	Fino a che punto il modulo di sintesi deve integrarsi con il sistema? Quali <i>features</i> devono essere sempre garantite dall'applicazione? . . . . .	2
2.7	É possibile fare maggior chiarezza sulla richiesta di un'implementazione separata delle varie componenti dell'applicazione? . . . . .	2
2.8	L'applicazione dedicata alla configurazione delle impostazioni del motore si può integrare nelle impostazioni di sistema di Android? . . . . .	3
2.9	É possibile velocizzare il tempo di campionamento della voce abbassando la durata del processo a un massimo di 10 minuti? . . . . .	3
2.10	Si riesce ad ottenere un buon risultato registrando nuove voci dal microfono di uno <i>smartphone<sub>G</sub></i> ? . . . . .	3
2.11	Quali strumenti di lavoro e di gestione ci verranno offerti? . . . . .	3

## 1 Informazioni Generali

- **Data:** 2016-03-09;
- **Luogo:** MIVOQ S.R.L., sede di Padova;
- **Ora:** 15:00;
- **Durata:** 90';
- **Partecipanti interni:** Stark Labs
  - Alberto Andriolo;
  - Francesco Bizzaro;
  - Enrico Chiara;
  - Riccardo Rizzo;
  - Federico Rossetto;
  - Gino Zaidan.
- **Partecipanti esterni:** MIVOQ S.R.L.
  - dott. Giulio Paci.

## 2 Questioni emerse

Di seguito sono trascritti gli argomenti che il gruppo ha trattato durante la riunione con la Proponente Mivoq. Il meeting ha avuto carattere informale.

A seguire, vengono riportate le domande poste accompagnate da relativa formalizzazione della risposta del referente dott. Giulio Paci.

### 2.1 Quale obiettivo si vuole raggiungere con l'applicazione?

Il capitolato è rimasto volutamente vago, per non imporre alcun limite sul tipo di applicazione. L'unico requisito primario è l'utilizzo del motore di sintesi, con lo scopo ultimo di metterne in risalto le potenzialità. Tuttavia è fortemente desiderabile che l'applicazione si dimostri utile per qualche utilizzo reale, in linea con gli esempi suggeriti nel capitolato.

### 2.2 Per quale motivo si è scelto il mondo mobile?

La scelta è stata dettata dal desiderio del committente di entrare in tale settore. La scelta è stata motivata, durante l'incontro, con le seguenti motivazioni:

- incentivare la diffusione dell'applicazione;
- praticità nell'avere l'applicativo su un dispositivo mobile, il quale risulta facilmente trasportabile per eventuali dimostrazioni della tecnologia FA-TTS<sub>G</sub>.

### 2.3 Si richiede che l'applicazione elabori una frase in tempo reale, senza usare audio creato precedentemente?

Possono essere fatte entrambe le cose, purché l'audio venga creato attraverso il motore di sintesi FA-TTS<sub>G</sub> con una connessione HTTP.

## 2.4 Supposto di utilizzare il motore di sintesi in un'applicazione di messaggistica, quali sarebbero i vantaggi offerti rispetto all'invio di note vocali?

Una nota vocale può essere solamente ascoltata, mentre attraverso la sintesi vocale un utente sarebbe libero di scegliere se leggere il messaggio o ascoltarlo sintetizzato, magari con la voce digitale del mittente. Di solito si preferisce leggere i messaggi, ma in alcune situazioni questo non è possibile (ad esempio se si sta guidando). A questo punto risulta utile avere la possibilità di farsi leggere il messaggio da un servizio TTS<sub>G</sub>.

## 2.5 Quali sono le principali problematiche riguardanti la connessione al servizio remoto?

Il servizio è offerto tramite un server interrogabile attraverso un'interfaccia HTTP<sub>G</sub>. Sono previste chiamate HTTP<sub>G</sub> in grado di fornire l'elenco delle voci disponibili, con le varie impostazioni modificabili.

Le problematiche possono essere di vario genere:

- la connessione può cadere;
- i tempi di risposta della connessione potrebbero non essere ottimali:
  - i file da trasferire sono codificati in formato WAV<sub>G</sub> e pertanto bisogna tenere conto anche del loro peso;
  - l'elaborazione richiede tempo.

Per poter arginare il problema del ritardo si potrebbero adottare strategie di *caching<sub>G</sub>*, conoscendo con anticipo i possibili testi da elaborare.

## 2.6 Fino a che punto il modulo di sintesi deve integrarsi con il sistema? Quali *features* devono essere sempre garantite dall'applicazione?

È desiderabile che venga realizzato un modulo software capace di integrare il servizio remoto offerto con il sistema mobile nel quale viene eseguita l'applicazione. Per risolvere la possibile assenza di connessione, che renderebbe impossibile lo sfruttamento del motore di sintesi, devono essere implementati meccanismi di *fallback<sub>G</sub>* che utilizzino le voci presenti nel sistema. In questo modo sarà possibile effettuare la sintesi con le voci di sistema e consentire all'applicazione di funzionare correttamente.

## 2.7 È possibile fare maggior chiarezza sulla richiesta di un'implementazione separata delle varie componenti dell'applicazione?

L'applicazione deve essere formata da 4 parti:

- un modulo per la sintesi, a sé stante, che corrisponde all'implementazione del motore FA-TTS<sub>G</sub> di Mivoq;
- un'applicazione per la configurazione, che deve essere in grado di interagire con il modulo per la sintesi al fine di modificarne determinati parametri. Ad esempio si vuole fornire la possibilità di aggiungere nuove voci, assieme a nuovi *preset<sub>G</sub>* di effetti associabili alle voci date;
- una libreria che faciliti l'utilizzo delle funzionalità aggiuntive, permettendo il riuso della libreria stessa;
- un'applicazione innovativa che dimostri in modo chiaro le potenzialità offerte dal suddetto modulo di sintesi.

## 2.8 L'applicazione dedicata alla configurazione delle impostazioni del motore si può integrare nelle impostazioni di sistema di Android?

È necessario che il gruppo approfondisca tale argomento autonomamente. Probabilmente non è possibile implementare tutte le funzionalità richieste direttamente nelle impostazioni del sistema operativo. Di conseguenza l'applicazione di configurazione dovrà svolgere questo compito esternamente. Tale applicazione dovrà consentire all'utente di aggiungere voci in maniera dinamica e di salvarle con i loro parametri all'interno del sistema. È quindi possibile sviluppare due applicazioni, oppure combinare le due funzionalità in un'unica applicazione.

## 2.9 È possibile velocizzare il tempo di campionamento della voce abbassando la durata del processo a un massimo di 10 minuti?

Attualmente il *competitor<sub>G</sub>* principale di Mivoq è circa 16 volte più lento rispetto ai 45 minuti richiesti dal motore FA-TTS<sub>G</sub> per registrare una nuova voce. Il motivo per cui il campionamento è così lungo deriva dalla necessità di raccogliere ed elaborare un elevato numero di dati, tant'è che solitamente occorrono più di 100 frasi per ottenere un buon risultato. Il sistema di registrazione offerto è ancora nelle fasi preliminari di sviluppo e al momento i tempi dichiarati sono i migliori disponibili.

## 2.10 Si riesce ad ottenere un buon risultato registrando nuove voci dal microfono di uno *smartphone<sub>G</sub>*?

La strumentazione degli *smartphone<sub>G</sub>* odierni è adatta ad ottenere un buon campionamento, il reale problema è dato dai tempi troppo elevati che interessano il processo.

## 2.11 Quali strumenti di lavoro e di gestione ci verranno offerti?

Mivoq fornisce il motore di sintesi assieme alle voci predefinite, materiale che può essere scaricato direttamente dal sito dell'azienda. Su richiesta del gruppo sarà possibile avere accesso ad un sito interno di Mivoq per campionare la propria voce e poterla riutilizzare anche all'interno dell'applicazione da realizzare. Non viene dato alcun vincolo sugli strumenti di sviluppo per implementare le applicazioni richieste. Sarà compito del team lo studio di eventuali *framework<sub>G</sub>* per facilitare il processo di sviluppo dell'applicazione finale.