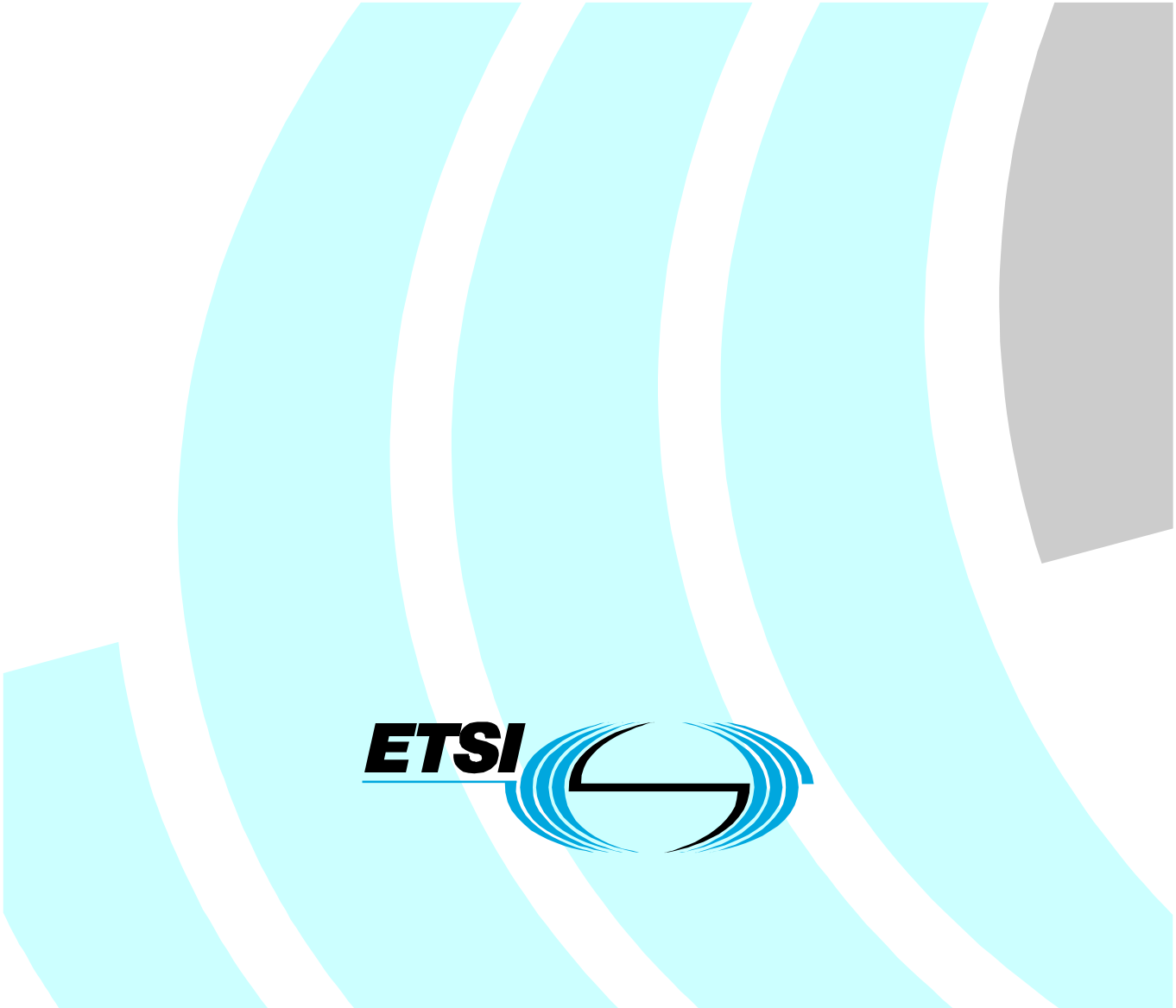
ETSI ES 202 076 V1.1.1 (2002-09)

ETSI Standard

Human Factors (HF);
User Interfaces;
Generic spoken command vocabulary
for ICT devices and services



Reference
DES/HF-00021

Keywords
ICT, interface, speech, telephony, user, voice

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Standard (ES) has been produced by ETSI Technical Committee Human Factors (HF).

The work has been conducted in collaboration with the industry. The present document is based upon expert knowledge, empirical data, user testing, acoustic discrimination tests and an industry-consultation and consensus process, aiming at a quick uptake and the widest possible support in product implementations to come.

Introduction

Telecommunication, converging with information processing, and intersecting with mobility and Internet, is leading to the development of new interactive applications and services, offering global access.

A technology enabling the most natural user interaction with these (often complex) systems and services is speech recognition. In recent years, speech recognition has become commercially viable on off-the-shelf ICT (Information and Communication Technology) devices and services. As the graphical user interface changed the way we interact with personal computers, voice user interfaces are changing communication.

Voice is a fundamental human paradigm for communications, forming an important foundation for universal access to the services and benefits of communications technology. Voice user interfaces are also a terminal, display and location-independent user interface technology, enabled by speech recognition technologies. In order to simplify the user's learning procedure and enable reuse of knowledge between different applications and devices, it is highly desirable to standardize the most common and generic navigation, command and editing vocabulary.

This effort is well timed, to catch the mass-market deployment of speech recognition enabled services, applications and terminals offering multi-lingual voice user interfaces. The well-defined, harmonized framework minimizes the learning curve, offering familiarity, knowledge transfer and developing user trust.

Uniformity in the basic interactive elements increases the transference of learning between devices and services using spoken commands and improves the overall usability of the entire interactive environment. Such transference becomes even more important in a world of ubiquitous devices and services using speech recognition.

The standardized, minimum generic set of spoken command vocabulary presented in the present document has been developed with a combined methodology, including tests with native speakers of the five languages (see annex A for details). Thereby, it provides useful help to developers of ICT devices and services, leading to quicker, more consistent, cheaper and better user interface development.

The work is aligned with and has been sponsored by the European Commission's initiative ***eEurope***, a program for inclusive deployment of new, important, consumer-oriented technologies, opening up global access to communications and other new technologies, for all (http://europa.eu.int/information_society/eeurope).

1 Scope

The present document specifies a minimum set of spoken commands required to control the generic and most common functions of ICT devices and services that use speaker independent speech recognition. It specifies the necessary and most common vocabularies to be supported by ICT devices and services for voice input, including command, control and editing.

The present document is applicable to the functions required for navigation, information retrieval, basic call handling and configuration of preferences. The present document also addresses the most common telecommunication services.

The present document specifies user tested commands for the languages with the largest number of native speakers in the European Union: English, French, German, Italian and Spanish, as spoken in their respective countries. Future revisions of the present document may include other languages, language versions and ICT commands (methodology guidance is provided in annex A).

The present document does not cover dialogue design issues, the full range of supplementary telecommunications services, performance related issues, natural spoken numbers covering more than one digit (other than "double") or speech output.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication and/or edition number or version number) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.

[1] ETSI ES 201 930: "Human Factors (HF); Specification of user requirements for use in ETSI deliverables".

[2] ETSI TR 102 068: "Human Factors (HF); Requirements for assistive technology devices in ICT".

[3] ETSI EG 202 116: "Human Factors (HF); Guidelines for ICT products and services; "Design for All"".

[4] ITU-T Recommendation E.161: "Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network".

[5] ETSI ETR 116: "Human Factors (HF); Human factors guidelines for ISDN Terminal equipment design".

[6] ETSI EG 202 048: "Human Factors (HF); Guidelines on the multimodality of icons, symbols and pictograms".

[7] ETSI EG 201 013: "Human Factors (HF); Definitions, abbreviations and symbols".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in EG 201 013 [7] and the following apply:

common command: command always available in any service, application or device for the user, under normal circumstances

communication impairment: difficulties in using speech, resulting from damage to structures involved, disorders of brain language processing, early childhood deafness, problems of muscular control, or co-ordination or other causes

device specific command: commands with a functionality that varies from one context to another

design for all: the design of products to be usable by all people, to the greatest extent possible, without the need for specialized adoption

dialogue: series of exchanges between the user and a system

function: the abstract concept of a particular piece of functionality in a device or service

ICT devices and services: devices or services for processing information and/or supporting communication, which has an interface to communicate with a user

impairment: any reduction or loss of psychological, physiological or anatomical function or structure of a user (environmental included)

keyword: word that the speech recognition system is looking for in word spotting mode (also known as "hot word" in voiceXML applications or "magic word" in GSM mobile telephony)

magic word: see **keyword**

menu: a menu offers a user a list of choices from which a selection can be made. A menu dialogue offers a user a series of lists of choices from which a series of selections can be made. The result from any one selection may be another menu

message: verbal or other auditory data recorded by users of a service. Callers, subscribers, or system administrators may record messages

spoken command: verbal or other auditory dialogue format which enables the user to input commands to control a device or service

supplementary service: additional service that modifies or supplements a basic telecommunication service

NOTE: Consequently, it cannot be offered to a customer as a stand-alone service; it has to be offered in association with a basic telecommunication service. The same supplementary service may be common to a number of basic telecommunication services. See ITU-T Recommendation I.210.

terminal: physical device which interfaces with a telecommunications network, and hence to a service provider, to enable access to a telecommunications service

NOTE: A terminal also provides an interface to the user to enable the interchange of control actions and information between the user and the terminal, network or service provider.

tinnitus: condition where noises are heard within the head or ear which can only be heard by the person affected

usability: the **effectiveness, efficiency** and **satisfaction** with which specified users can achieve specified goals (tasks) in a particular environment, see ETR 09

NOTE: In telecommunications, usability should also include the concepts of learnability and flexibility; and reference to the interaction of more than one user (the A and B parties) with each other and with the terminals and the telecommunications system, see ETR 116 [5].

user: the person who uses a telecommunications terminal to gain access to and control of a telecommunications service

NOTE: The user may or may not be the person who has subscribed to the provision of the service. Also, the user may or may not be a person with an impairment, e.g. elderly or disabled persons.

user interface: the physical interface through which a user communicates with a telecommunications terminal or via a terminal to a telecommunications service

NOTE: The communication is bi-directional in real time and the interface includes both control and display elements.

user requirements: requirements made by users, based on their needs and capabilities, on a telecommunication service (e.g. the UPT service) and any of its supporting components, terminals and interfaces, in order to make use of this service in the easiest, safest, most efficient and most secure way

word spotting mode: a special state of the recognition system in which no speech is recognized or processed other than a limited set of key words

NOTE: A typical usage is in a dormant state of the speech recognizer, where issuing a "wake up" command (also known as hot-word or key-word) can reactivate speech functionality.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

*	the Star on a standard telephone keypad array, see ITU-T Recommendation E.161 [4]. Also known as the Asterisk key
#	the Hash on a standard telephone keypad array, see ITU-T Recommendation E.161 [4]. Also known as the Square, Sharp, or Number sign ("pound" in the USA)

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ASR	Automatic Speech Recognition
ETSI	European Telecommunications Standards Institute
GSM	Global System for Mobile telecommunication
ICT	Information and (tele) Communication Technologies
ITU-T	International Telecommunications Union - Telecommunication Standardization Sector
MMI	Man-Machine Interface
UPT	Universal Personal Telecommunication

4 User requirements

Intended *users* of the present document are those deploying and implementing it, the interaction designers and other developers of ICT devices and services with a speech user interface.

Intended *end users* mentioned in the present document are consumers of ICT devices and services, ranging from first time to experienced power users, who can produce intelligible, natural speech.

The end user's main goal is to use ICT devices and services efficiently under their intended circumstances.

Uniformity in the basic interactive elements increases the transference of learning between devices and services using spoken commands and improves the overall usability of the entire interactive environment. Such transference becomes even more important in a world of ubiquitous devices and services using speech recognition. Therefore, deployment of the present document will enable users to reapply knowledge and previous experience between different devices and services and to control common functions by using a generic vocabulary of spoken commands.

Certain end users with special needs, such as very young children who do not read yet, the visually disabled, users with inability to perceive tactile stimuli, with temporary alterations to the sense of touch or limited dexterity in the hands, will benefit from a generic spoken command vocabulary.

Other end users with special needs might experience difficulties trying to use speech recognition based user interfaces. Generic commands or not, the present document will not make a difference in the case of temporary or permanent difficulties caused by communication impairments such as: speech impairments, cognitive problems or the lack of necessary level of proficiency in the respective language, and in the case of dialogues with spoken responses, the inability to hear sounds at the usual volume and frequency, tinnitus, difficulty in distinguishing and understanding speech or temporary hearing difficulties.

The spoken commands must have the following usability characteristics:

- Easy to learn
- Memorable
- Natural
- Unambiguous

A well-designed speech user interface of an ICT device or service should have:

- A shallow learning curve
- Generic (standardized) ways to accomplish most common tasks
- The ability to handle the vagaries of speech recognizers in a reliable and predictable way, maximizing the user experience.

The ETSI standard form in ES 201 930 [1], "Form describing User Requirements in ETSI deliverables", has been used as a reference while specifying the user requirements. The methodology for collecting and validating spoken commands is described in annex A.

For further guidance, including specifics of user impairments and resulting handicaps, assistive technologies, design for all and multi-modal interfaces, see TR 102 068 [2], EG 202 116 [3] and EG 202 048 [6].

5 List of commands

5.1 Principles of use

The spoken commands in the present document are divided into two major categories:

1) Common commands (always available in any device or service for the user) and

2) Domain specific commands (with a functionality that differs from one context to another).

For the entire ES, the following three principles of use in implementations apply, assuming a speech recognition user interface is provided:

1) The ICT device or service must support all the common user commands specified in the present document. If a common command cannot be supported by the ICT device or service (e.g. due to lack of the specific functionality), the common command must still be accepted as user input and guidance information provided to the user.

2) For domain-specific functionality supported by the ICT device or service, domain specific user commands must be supported.

3) In addition to the commands specified in the present document, alternative and additional commands may be offered.

5.2 Common commands

Common commands are commands always available in any service, application or device for the user, under normal circumstances. There are two exceptions to this, where common commands are not necessarily available:

- in interaction contexts, where a limited number of spoken commands might be enforced due to application design specifics (e.g. where global commands might interfere with a confirmation in a financial transaction); and
- in word spotting contexts (e.g. during an ongoing phone call), where only hot/key/magic-words are available

Common commands are divided into two sub-groups:

1) Context independent common commands; and
2) Context dependent common commands.

5.2.1 Context independent common commands

The context independent common commands in table 1 maintain the same functionality, irrespective of the (dialogue) context in which the commands are executed.

Table 1: Context independent common commands

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
1.1	List commands and/or functions	**Options**	**Choix**	**Menü**	**Menù, Lista comandi**	**Opciones**	Request for listing of available words (optionally with their functionality)
1.2	Terminate service	**Goodbye**	**Quitter, Au revoir**	**Benden**	**Spegni, Fine**	**Salir**	End call, get off line, end session
1.3	Go to top level of service	**Main menu**	**Menu principal**	**Haupt-menü**	**Menù principale**	**Inicio, Menú principal**	Leave sub-application, go to main menu or application
1.4	Enter idle mode	**Standby**	**Veille**	**Stand-by**	**Sospendi, Stand-by**	**Espera**	Put the Automatic Speech Recognition (ASR) into monitoring mode for a wake-up command
1.5 (See note)	Transfer to human operator	**Operator**	**Assistance**	**Hotline, Service**	**Operatore, Assistenza**	**Operador**	Leave the speech recognition mode and transfer to a human attendant, an operator, in telecommunications-specific contexts. This command should also be used when offering relay services
1.6	Go back to previous node or menu	**Go back**	**Retour**	**Zurück**	**Indietro**	**Atrás**	Navigate backwards in a dialogue structure (can also be used to cancel a forced choice operation)
NOTE: The command "Helpdesk" would be recommended for IT-specific contexts. However, it conflicts with the common command 2.1 "Help" in several languages and causes recognition conflicts from the ASR point of view.							

5.2.2 Context dependent common commands

The context dependent common commands in table 2 can have a functionality that might differ from one context to another.

NOTE: The function itself is context dependent, not the effect of the operation.

Table 2: Context dependent common commands

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
2.1	Help	**Help**	**Aide**	**Hilfe**	**Aiuto**	**Ayuda**	Provide context dependent explanations and guidance (may provide more detailed help on repetition of the command)
2.2	Read prompt again	**Repeat**	**Répéter**	**Wiederholen**	**Ripeti**	**Repetir**	Repetition of the last acoustic feedback message

5.3 Domain (application, device and/or service) specific commands

Domain specific commands might have a functionality that varies from one context to another. The commands in this clause may be used in any device or service having the relevant functionality.

5.3.1 Core commands

The core commands in table 3 are general commands that have not been categorized in other clauses.

Table 3: Core commands

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
3.1	Confirm operation	**Yes**	**OK, Oui**	**Ja, OK**	**Sì, Confermo**	**Sí**	Positive confirmation
3.2	Reject operation	**No**	**Non**	**Nein**	**Annulla, No**	**No**	Negative confirmation
3.3	Cancel current operation	**Stop**	**Stop**	**Stopp, Abbruch**	**Stop, Interrrompi**	**Cancelar**	Immediately abort ongoing operation (e.g. during the (long) playback of a recorded message)

5.3.2 Digits

The commands in table 4 apply to the entering of digits and telephone numbers or telecommunication services.

Natural, spoken numbers representing more than one digit other than "double", are outside the scope of the present document.

Table 4: Digits

Index (see note 1)	**ICT device/service function**	**English spoken command**	**French spoken command**	**German spoken command**	**Italian spoken command**	**Spanish spoken command**	**Explanation**
4.1	Enter digit 0	**Zero, Oh**	**Zéro**	**Null**	**Zero**	**Cero**	Enter the digit "zero"
4.2	Enter digit 1, 2, 3, 4, 5, 6, 7, 8 or 9	**One, Two, Three, Four, Five, Six, Seven, Eight, Nine**	**Un, Une, Deux, Trois, Quatre, Cinq, Six, Sept, Huit, Neuf**	**Eins, zwei, zwo, drei, vier, fünf, sechs, sieben, acht, neun**	**Uno, Due,Tre, Quattro, Cinque, Sei, Sette, Otto, Nove**	**Uno, Dos, Tres, Cuatro, Cinco, Seis, Siete, Ocho, Nueve**	Enter the digits "one" to "nine"
4.3 (See note 2)	Indication that the next digit is repeated twice	**Double**	**Double**	**----------**	**Doppio**	**Doble**	Enter the next digit twice
4.4	Enter international access code	**International, Plus**	**International, Plus**	**Plus, Null-null**	**Zero zero, Internazionale, Più**	**Internacional, Más**	The command will enable placing calls using the standard international number format
NOTE 1: Although # (Hash, Pound, Square) and * (Star) are frequently used and typical telecommunication service delimiters, they are not standardized because these are typically keypad-interaction oriented, non-verbal commands. NOTE 2: The command word for Double is not used in the German language.							

5.3.3 Name and digit dialling

The commands in table 5 apply to the retrieval of available telephone numbers and call set-up.

Table 5: Name and digit dialling

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
5.1 (see note 1)	Initiate digit dialling sequence	**Dial**	**Composer**	**Wählen**	**Componi**	**Marcar**	Initiate a call to a number
5.2 (see note 1)	Dial a number or name	**Call**	**Appeler**	**Verbinden mit**	**Chiama**	**Llamar**	Initiate a call to a name or number
5.3	Home phone number (location)	**Home**	**Maison**	**Privat**	**Casa**	**Casa**	Call the stored home number
5.4	Work phone number (location)	**Work**	**Travail, Bureau**	**Büro, Arbeit**	**Ufficio, Lavoro**	**Trabajo**	Call the stored work number
5.5	Mobile phone number (location)	**Mobile**	**Mobile, Portable**	**Mobil, Handy**	**Cellulare**	**Móvil**	Call the stored mobile number
5.6	Car phone number (location)	**Car**	**Voiture**	**Auto**	**Auto**	**Coche**	Call the stored car number
5.7	Personal number (attribute)	**Personal number**	**Numéro personnel**	**Eigene Nummer**	**Numero personale**	**Número personal**	As above, but specifying the person's number which redirects all calls to their chosen number
5.8 See note 2	Make a call to the emergency services	**Emergency**	**Urgences**	**Notruf**	**Emergenza, SOS, Soccorso**	**Emergencias**	Call the emergency service centre See note 1

NOTE 1: There is no need to explicitly support the commands Call and Dial if, for example, these are activated automatically in the ICT device or service.

NOTE 2: The spoken command "Emergency" shall be supported. However, as a variety of numbers are used on national levels, additional names (e.g. Ambulance, Police) for calling national emergency services should be included, as these command might be used in a serious emergency situation (numbers, e.g. 112, 999- will be recognized anyway, if digit recognition is available). This is of even higher importance in cases where manual dialling is not available.

5.3.4 Basic call handling and supplementary services

The commands in table 6 are applicable to basic call set-up and the interaction needed to use supplementary telecommunication services.

Activation of some services might require word-spotting technologies to be available and activated in order to recognize a wake-up command (e.g. handling of an incoming call, already having one connection active).

Table 6: Basic call handling and supplementary services

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
6.1 (see note)	Accept incoming call	**Answer**	**Allô, Répondre**	**Gespräch annehmen, Ja**	**Rispondi, Sì**	**Contestar**	Accept incoming call
6.2 (see note)	Reject incoming call	**Busy**	**Occupé**	**Abweisen, Nein**	**Occupato, No**	**Ocupado**	Do not accept incoming call
6.3 (see note)	Forward incoming call	**Divert to**	**Transférer**	**Umleiten zu, Weiterleiten an**	**Inoltra**	**Desviar**	Instead of taking the incoming call, send it on to another number
6.4	Redial last dialled number	**Redial**	**Rappeler**	**Wahlwiederholung**	**Richiama, Ripeti numero**	**Rellamada**	Dial the last dialled number once again
6.5	Set up a call-back to a called number	**Keep trying**	**Rappel automatique, Insister**	**Rückruf**	**Riprova**	**Continuar llamando**	Call completion on no reply or busy (also known as Call-back)
6.6	Set up a conference call	**Conference call**	**Conférence**	**Konferenz**	**Conferenza**	**Llamada múltiple**	Connect a minimum of two phone numbers to a single phone call
6.7	Set up a call diversion	**Divert all calls to**	**Transférer appels au**	**Alle Anrufe weiterleiten an**	**Trasferisci chiamata a**	**Desviar todas las llamadas**	Redirect all future incoming calls to a specified number
6.8	Transfer an ongoing call	**Transfer**	**Transmettre**	**Weiterleiten an**	**Trasferisci a**	**Transferir**	During a call, transfer the other party to a third number
6.9	Put call on hold	**Hold**	**Attente**	**Halten**	**Attesa**	**Mantener**	Park an ongoing call
6.10	Switch between two calls (hook flash)	**Switch call**	**Basculer**	**Wechseln zu, Makeln**	**Passa, Cambia**	**Llamada en espera**	Resume the call on the held line
NOTE: These commands (6.1, 6.2 and 6.3) shall only be used in incoming call situations.							

5.3.5 Media control

The commands in table 7 apply to the control of audio and video, covering basic multi-media functionality (e.g. recorded audio and video).

Table 7: Media control

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
7.1	Play a recording	**Play**	**Lire**	**Wiedergabe**	**Ascolta**	**Reproducir**	Retrieve a recording
7.2	Stop temporarily	**Pause**	**Pause**	**Pause**	**Pausa**	**Pausa**	End playing a recording (for the moment)
7.3	Resume interrupted playback	**Continue, Play**	**Reprendre**	**Weiter**	**Continua**	**Continuar**	Continue playing a recording
7.4	Stop playing a recording	**Stop**	**Stop**	**Stopp**	**Stop**	**Parar**	Stop
7.5	Move forward faster than play	**Fast forward**	**Avancer**	**Vorspulen**	**Avanti veloce**	**Adelante rápido**	Go forward faster in a recording
7.6	Move backward	**Rewind**	**Rembobiner**	**Zurückspulen**	**Indietro**	**Rebobinar**	Go backward in a recording

5.3.6 Browseable list for navigation

The commands in table 8 apply to navigation and search in lists of various kinds.

Table 8: Browseable list for navigation

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
8.1	Go to next item	**Next**	**Suivant, Suivante**	**Weiter**	**Avanti, Successivo**	**Siguiente**	Go to the next item
8.2	Go to previous item	**Previous**	**Précédent, Précédente**	**Zurück**	**Precedente**	**Anterior**	Go back to the previous item
8.3	Provide more information about selected item	**Details**	**Détails**	**Details**	**Dettagli**	**Más información, Abrir**	Give information about the attributes of a selected item

5.3.7 Editing commands

The commands in table 9 apply to the editing of text, audio and video.

Table 9: Editing commands

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
9.1	Modify item	**Edit**	**Modifier**	**Ändern**	**Modifica**	**Modificar**	Select an item for modification
9.2	Remove item	**Delete**	**Supprimer**	**Löschen**	**Elimina**	**Borrar**	Delete an item
9.3	Store item	**Save**	**Sauvegarder**	**Speichern**	**Salva**	**Guardar**	Save an item
9.4	Respond to item	**Reply**	**Répondre**	**Antworten**	**Rispondi**	**Responder**	Draft a reply
9.5	Forward item	**Forward**	**Faire suivre**	**Weiterleiten**	**Inoltra**	**Reenviar**	Forward an item
9.6	Create new item	**Add**	**Ajouter**	**Neu, Hinzufügen**	**Aggiungi**	**Añadir**	Add new item
9.7	Send item	**Send**	**Envoyer**	**Abschicken**	**Invia**	**Enviar**	Transmission of any pre-prepared data (e.g. e-mail)
9.8	Move item to a new location	**Move**	**Déplacer**	**Verschieben**	**Sposta**	**Mover**	Transmission of stored data to a named location
9.9	Start a recording	**Record**	**Enregistrer**	**Aufnahme**	**Registra**	**Grabar**	Typically, memo and mail applications
9.10	Play a recording	**Play**	**Marche**	**Abspielen**	**Ascolta**	**Reproducir**	Retrieve a recording
9.11	Pause playback or recording	**Pause**	**Pause**	**Pause**	**Pausa**	**Pausa**	End playing a recording (for the moment)
9.12	Resume interrupted playback or recording	**Continue**	**Reprendre**	**Weiter**	**Continua**	**Continuar**	Continue playing a recording
9.13	Reverse the previous action	**Cancel**	**Annuler**	**Abbrechen, Abbruch**	**Cancella, Indietro**	**Cancelar**	Go back one step in the history of states (repeating the command should go back one step further in the history
9.14	Reapply the undone action	**Redo**	**Refaire**	**Wiederherstellen**	**Ripeti**	**Rehacer**	Go forward one step in the history of states (repeating the command should go forward one step further in the history)

5.3.8 Device settings

The commands in table 10 apply to device settings. These commands do not apply to network based services.

Table 10: Device settings

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
10.1	List networks	**Choose network**	**Choisir le réseau**	**Netz wählen**	**Scegli rete**	**Selección de red**	List currently available networks
10.2	Increase the volume	**Volume up, Louder**	**Augmenter volume, Plus fort**	**Lauter**	**Alza volume, Aumenta volume**	**Subir volumen**	Increase voice output level
10.3	Decrease the volume	**Volume down, Quieter**	**Diminuer volume, Moins fort**	**Leiser**	**Abbassa volume**	**Bajar volumen**	Decrease voice output level
10.4	Silent mode	**Sound off**	**Couper son**	**Ton aus**	**Suono off, Suono non attivo**	**Apagar sonido, Desactivar sonido**	Mute ring signals and tones
10.5	Re-activate the audio output	**Sound on**	**Remettre son**	**Ton an**	**Suono on, Suono attivo**	**Encender sonido, Activar sonido**	Activate all audio output
10.6	Silence the loudspeaker	**Speaker off**	**Couper haut-parleur**	**Lautsprecher aus**	**Altoparlante off, Altoparlante non attivo**	**Apagar altavoz Desactivar altavoz**	Only the loudspeaker is muted
10.7	Reactivate the loudspeaker	**Speaker on**	**Activer haut-parleur**	**Lautsprecher an**	**Altoparlante on, Altoparlante attivo**	**Encender altavoz, Activar altavoz**	Only the loudspeaker is activated
10.8	Re-activate the microphone	**Mike on**	**Activer micro**	**Microfon an**	**Microfono on, Microfono attivo**	**Encender micrófono, Activar micrófono**	Only the microphone is activated
10.9	Silence the microphone	**Mike off**	**Couper micro**	**Microfon aus**	**Microfono off, Microfono non attivo**	**Apagar micrófono, Desactivar micrófono**	Only the microphone is muted
10.10	Activate vibrating alert	**Vibrate on**	**Activer vibreur**	**Vibrationsalarm an**	**Vibrazione on, Vibrazione attiva**	**Encender vibración, Activar vibración**	Causes the device to vibrate (instead of, or in addition to, giving an audible signal)
10.11	Deactivate vibrating alert	**Vibrate off**	**Désactiver vibreur**	**Vibrationsalarm aus**	**Vibrazione off, Vibrazione non attiva**	**Apagar vibración, Desactivar vibración**	Switches off the vibrating alert
10.12	Summary of current device status	**Status**	**État**	**Status**	**Stato**	**Estado**	Typically, indicates battery status and network coverage details (e.g. time and date)
10.13	Change profile (pre-stored settings)	**Profile**	**Profil**	**Profile**	**Profili**	**Perfil**	For example, one profile for noisy environments and one for quiet environments (or when a user does not want to be disturbed by calls)

5.3.9 Word spotting mode

The commands in table 11 apply to the activation of the speech recognizer in ICT devices and services.

Table 11: Word spotting mode

Index	ICT device/service function	English spoken command	French spoken command	German spoken command	Italian spoken command	Spanish spoken command	Explanation
11.1 (see note)	Wake-up the speech recognizer (ICT device or service in word spotting mode)	**Wake-up**	**Activer**	**Aktivieren, Start**	**Riprendi, Attiva**	**Activar**	ASR ignores all speech input, except a wake-up command (hot-word, magic word or keyword). When this command is detected, the recognizer switches to a larger active vocabulary, determined by the dialogue design
NOTE: In addition to this command, the ICT device or service may use the device or service name, as a synonymous command.							

Annex A (informative): Methodology for defining command vocabularies

In this informative annex, we describe how the vocabulary words defined in the present document have been determined. This should also constitute a useful guideline for extending the vocabularies to other languages. The annex A first describes the general methodology, with various different implementations of experiments, and later concentrates on the actual path chosen in the making the present document.

A.1 Introduction

The two main perspectives of the methodology for collecting and validating spoken commands are:

- for users, the command words are both intuitive and easy to remember, while
- a speech recognition system requires commands to be easily discriminable.

In the scientific literature, several methods have been described by authors such as Book, Goldstein, Guzman and MacDermid which we have used in preparing the present document.

The methodology consists of several distinct steps:

1) **Spontaneous generation of potential command words**. The purpose of this step is to make an inventory of words that humans would intuitively use, given the task that they want to complete.

2) **Confidence rating of the found potential command words**. The purpose of this step is to ensure that the words found in the first step are considered likely to complete the task when a test subject is given the choice to use the word.

3) **Phonetic discrimination**. The purpose of this step is to ensure that command words that can be active simultaneously in a dialogue context can be recognized correctly by the speech recognition system.

The literature also describes a *recall test*, to be performed between steps 2 and 3. The purpose of this test is to verify that command words that have been learnt by test subjects can be remembered easily some time later. However, in this annex, we will not consider this test.

There are several ways of performing each step. In the following clauses we will explain the methodologies in further detail.

A.2 Spontaneous generation of command words

In order to ensure that command words for speech recognition enabled devices and services are intuitive, some evidence must be gained as to which word(s) a user would use without prior training or experience. It is not trivial to find this out, because in order to get this kind of information the (potential) user is likely to be primed for certain words or phrases. For instance, if a test is set up where test subjects are explained the task to be performed it is likely that the explanation contains some words that are candidate command words. If, on the other hand, dialogues of users of actual running systems are analysed it is likely that the command words found in these dialogues are the words that the system designer has chosen and that the user has learnt to use.

Here we describe two methods that allow the collection of spontaneous command words from test subjects. In both methods, test subjects play a vital role. They must be recruited among people who understand the services for which the command words are sought, and are familiar with the functionality, but are not actual users of speech-enabled implementations of such services.

First, for all services, the conceptual functionality to be supported must be determined and described (indicated in the second column of the various tables in clause 5 of the present document).

Second, for each of the functionalities the test subject must be explained which functionality is meant, without priming the test subject for particular words.

A.2.1 Storyboard method

In this method a professional artist makes a so-called storyboard, a set of illustrations or cartoons, for each function. The test subject is explained the background and shown the illustration (an example is shown in figure A.1), and is asked to say the command she would use in order to activate the shown functionality. This method has been described by MacDermid and Goldstein (1996).

The advantage of this method is that the same storyboard can be used for several different languages, as long as there are limited cultural constraints involved. The disadvantage is that some functionality might be very difficult to describe pictorially, and that there can be quite a lot of effort necessary from the artist.

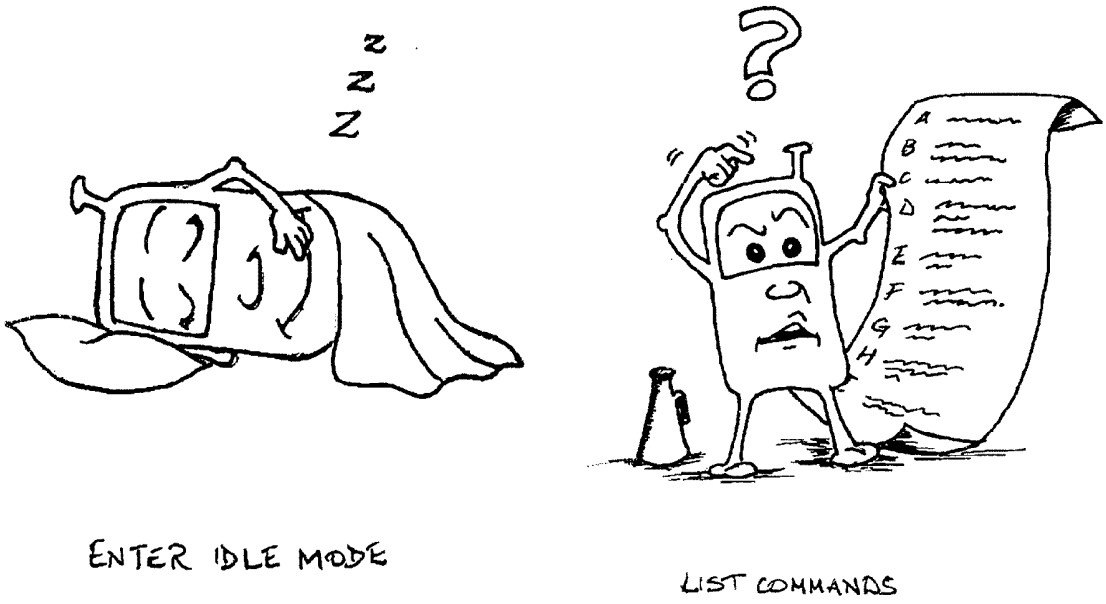


Figure A.1: Examples of single picture storyboards, for the commands "list commands" and "enter idle mode"

A.2.2 Carefully worded descriptions method

In this method the functionalities are described textually in a paragraph of text, which is carefully constructed not to use any word, which might possibly be used as a command word. This method has been described by Guzman.

The advantage of this method is that there is no need for a highly skilled professional for generating the textual descriptions. The disadvantages are that the descriptions may sometimes turn out to be very clumsily constructed in order to prevent using an obvious command word, and that this effort must be carried out in all languages one wants to conduct the inventory in. Also, these must be carefully developed for each target language.

A.3 Confidence rating of command words

When the spontaneous words have been generated with a sufficient number of test subjects, a histogram of the word frequencies can be made. This histogram gives a lot of information about the variability in responses. Thus, it can be seen immediately what the most likely command words are. Integration of the sorted frequencies indicates what the coverage of spontaneous commands is, if the most frequent words are available for recognition.

For the confidence-rating test described, one can select the most frequent words that cover at least a given percentage (say, 80 %) of the spontaneously generated words. For instance, for "confirmation" this might be "yes," "sure" and "no problem," if these are the most frequently generated commands and cover 85 % or the responses for the functionality "confirmation".

The procedure described above does not guarantee that the command words imply the targeted functionality. For instance, for functionality "confirmation" the command phrase "why not" might have come up in the list of spontaneous commands, but reversibly it might not be obvious to a user that the command "why not" implies confirmation; it might suggest the inquiry of a reason. A confidence rating of command words tests how likely it is that the given command word implies the correct functionality.

Because the "spontaneous generation of words" test is an open response experiment, many different expressions for very similar command words can be obtained. Therefore, a manual check of the histogram can be necessary, where responses with the same the essential term are grouped together. Thus, for each function, a set of candidates for the confidence test can be obtained.

A way of measuring the confidence of command words is the following, after the work of Guzman. A group of test subjects that are independent of the ones used to generate command words is presented the same data as in the first test, but are requested a different response. The test data can again be either from the "storyboard" or "descriptions" method. Instead of asking for a spontaneous command, the candidate commands from the first test are shown. For each of the commands, the test subject are asked how confident they are that the command word will imply the functionality, on a 5-point scale.

Instead of an explicit confidence measure, the subjects can be asked to choose the command word for which they have most confidence.

A.4 Phonetic discrimination

For a voice-enabled application it is essential that command words are recognized correctly. Although a system can ask for confirmation for certain not undoable operations (e.g. "delete subscription"), it is not acceptable if almost every command needs confirmation (e.g. "do you want to hear the next item? Please say yes or no"). For a given application or service there will be several contexts defined in which certain command words will be available. E.g. in the context of "confirmation question," words like "yes," and "no" will be active in the ASR vocabulary. The number of incorrectly recognized commands can be reduced if the available words in a given context are acoustically reasonably different.

There are several ways to test the discriminability. We will assume that for the service or application the ASR contexts are well defined. For a given context, there will be a number of words active.

A.4.1 Recognizer field test

One can find out the acoustic discriminability by a field test with a real speech recognition system. This test gives realistic discriminability measures, but the results are sensitive to many chosen parameter setting. Some of these are

- The type of recognition system (brand, speaker dependency, noise robustness, etc.)
- The test database (recorded speech samples versus live speech from test subjects, speaking style, etc.)

The recording of test databases for voice commands requires quite a lot of effort, but there are several projects such as SPEECON, Dialog 2000 and SpeechDat-Car- in which databases for voice commands are gathered. The test is conducted by preparing the ASR to recognize the required command set for each context, and then test each context with several instances of all the available commands within the context, uttered by many different test subjects. The *confusability* of a command word *A* with respect to an alternative command word *B* can be defined as the fraction of times and utterance of word *A* is recognized as word *B* by the recognition system. A confusion matrix for each context, containing the confusability of all active menu words with respect to each other, can indicate which command words pose particular problems to the recognition system. The *discriminability* of a set of command words is a measure that characterizes the whole confusion matrix.

If the test database consists of recorded speech, the test can be repeated for another ASR system. This will give insight in the recognition system dependency of the discriminability results.

A.4.2 Pronunciation dictionary test

An alternative to a field test is the analysis of the acoustic realizations of the command words. This can be performed without collecting speech databases or test subjects, but the predictions are not validated. The only thing necessary is a *pronunciation dictionary*, a tool that is used often by speech recognition ICT device or service developers. A pronunciation dictionary consists of a lookup table of words in terms of their phone (a separate unit of sound, similar to a phoneme in linguistics) sequences. For instance, the phone sequence for "yes" may be specified as the sequences "j eh s" or "j ea" (where we have introduced a Latin character readable phone symbols "j" "eh" "ea" and "s"). Typically for a Western language phone sets of 40-60 phones are defined for ASR systems.

The acoustic discriminability of two command words can be predicted on the basis of the phone sequences of the words. The number of different phones (order is important) might be called the first order prediction of the discriminability. For instance, in the context "start" "stop" the number of different phones is 2 for "stop" and 3 for "start". This is a relatively low number compared to the number of phones in the words, respectively 4 and 5. As a contrast, the context "begin" "end" has no phones/positions in common, so the number of different phones is 5 and 3, respectively.

A more elaborate scheme takes into account the confusion probability of two phones: e.g. most ASR systems (as well as humans) have difficulties always distinguishing between "m" and "n". For a particular ASR system these phone confusion probabilities may be measured, but this requires a quite elaborate test set-up of the ASR system. If this information is not available, the phones in a language might be grouped, and the discriminability can be measured in terms of the different phone groups. For example, if "p" and "t" are in the same phone group (plosives), the words "top" and "pot" have all phone group/positions in common, and the predicted discriminability is very low.

For some languages pronunciation dictionaries are publicly available. However, the complete set of command words in a service or product is limited, and the individual pronunciation of the command words can be found by consulting an expert phonetician. This person can also help in specifying groups of phones that can be considered "very similar".

A.4.3 Applying the acoustic discrimination

The discriminability measure can be used to find the optimally performing command words for each recognition context, a complex procedure. An example can help to clarify the procedural difficulties. Suppose, for instance, that in a media browsing application the functions "move to first message" and "exit application" are available simultaneously. Suppose further that for the first function the commands "top" and "first" come out of the confidence test with preferences 70 % and 30 %, while for the second function the commands "quit" and "stop" appear to have preference levels 25 % and 75 %, respectively. Without paying attention to acoustic discrimination, the command words "top" and "stop" would be the preferred ones. If the acoustic discriminability is taken into account, however, which word is going to be replaced by an alternative command with lower subjective preference? The percentages for the alternative words, 30 % and 25 % respectively, appear very similar, and moreover they may not be the only important factor. There might be other words for which discriminability plays a role (e.g. a command word "quick" with high subjective preference). This means that discriminability optimization is process that should be applied to the whole menu structure, possibly involving different contexts and even different applications.

It is quite difficult to formally define a procedure for optimizing the command vocabulary words, because many more factors should then be incorporated such as frequency of occurrence of the commands and likelihood that other applications will be available. A more pragmatic approach to the problem therefore is the following procedure:

a) For each context, start with the command words suggested by the confidence rating test.

b) Find possible pairs of commands that give rise to acoustic discriminability problems.

c) Choose an alternative for one of the command words, with minimum repercussion with respect confidence rating.

d) Repeat step b) and verify that there are no other commands that clash acoustically with the alternative command word.

e) Repeat step a) to verify that all functions that have new alternative commands do not occur in other contexts or have no acoustic discriminability problems there.

This procedure assumes that there is a relatively low probability that two command words will have low acoustic discriminability.

A.5 Application of the methodology to the present document

In this clause, the work methodology applied, developing the present document, is presented. The specific choices were made on the basis of experience, available resources and expert opinion.

A.5.1 Identification, definition and selection of application areas and key functionality

In the very early phase of the work, key areas of typical ICT device and service functionality were defined, collected, listed and evaluated. Belonging commands were considered, grouped, categorized and reduced to a generic, minimum sub-set of functionality and belonging commands.

The considered input consisted of empirical data, off-the-shelf products, expert knowledge and previous work mentioned in annex B: Bibliography.

A.5.2 Spontaneous command generation test

This test was carried out through an interactive web survey for all considered languages. Subjects were acquired from various sources and either paid for their work or compensated for their efforts in another way (e.g. entering a lottery with mobile phones and wine offered as prizes to be won). The survey was placed on an unpublished web site so that there was control over who took part in the survey. Subjects not having web-access of their own, were provided web access by the test leader.

For the functions in clause 5, carefully worded descriptions were generated. For all of the languages considered in the present document, the descriptions were reviewed by an expert, knowledgeable in all of the languages, in order to ensure consistency across the languages. For each description, the subjects were asked to provide a command word or phrase. Optionally, they could give one alternative word or phrase.

The form was submitted to a central processing facility that collected all the subjective data in a data base. Responses were first normalized in spelling and responses with the same essence were grouped under the essential word. For each of the commands a frequency histogram was made, weighing the first choice twice as much as the (optional) second choice. Then the two to six commands that contributed to the majority of the responses were selected for the confidence test.

A.5.3 Confidence rating ("Multiple Choice") test

The confidence was measured by a forced Multiple Choice test. Like the spontaneous generation experiment, the data was collected through an interactive web survey. Subjects were acquired along similar lines and it was verified that the subjects had not participated in the first survey.

The presented commands have been reviewed by the experts and in some cases, complemented with other candidate commands in cases where strong evidence suggested these should be included.

For each of the functions in the Multiple Choice test, statistics of the responses were produced. This information was used to perform the acoustical discrimination verification procedure.

The total number of test subjects used in the spontaneous command generation and confidence rating "Multiple Choice" tests was 329.

A.5.4 Acoustical discriminability

The acoustical discriminability was carried out along the lines of the procedure indicated in clause A.4.3. In order to determine acoustical similarity between two command words the pronunciation of standard pronunciation lexicons were used. Similarity of was based on grouping of phones in similar acoustic-phonetic classes.

A.5.5 Final decisions

The final choice of command vocabulary for each language was based on a joint expert opinion, considering a common weight of the following aspects:

1) The confidence test rating.

2) Evidence in literature.

3) Personal experience.

4) The acoustical discriminability, based both on pronunciation and ASR field experience.

A.6 Example of data collection

As an example, the procedure for determining the command word for function 1.1, "List commands and/or functions", is given below.

In the web survey, subjects were asked what command word or phrase they would like to say in the following situation:

Command 1 The ICT device or service is waiting for you to say a command but you do not know which commands are available to you

Respondents could give two alternative commands for each description. The first suggestion was given a score of 1 and the second suggestion scored ½. Then the total score for each suggestion was calculated, combining suggestions that only have small variations on inspection. Thus, for the first description, the frequencies tabulated in table A.1 were obtained:

Table A.1: Raw response histogram data for the command "List all commands and/or functions" (words that have been Italicized are later counted as either "options" or "commands")

First choice		Alternative choice (weight ½)	
Frequency	**Responded command**	**Frequency**	**Responded command**
6	*Options*	2	Menu
5	Menu	2	Help
3	List commands	1	Hi
2	Help	1	What *options* are available
2	Hello	1	What can I do
1	What *options* do I have	1	Ready
1	What can I do	1	*Options*
1	Wait query	1	Menu please
1	Starting up *commands*	1	Main menu
1	*Options* please	1	List
1	List *options*	1	Identify *commands*
		1	Hold—what *command*
		1	*Commands* help
		1	*Commands*
		1	*Commands* menu
		1	Available *commands*

Analyses showed that both the command words "options" and "command" occurred as essential word in various phrases in the tail of the histogram. These phrases have been grouped under the essential word, leading to the most frequently used words for the Multiple Choice test: Options (9), List commands (7), Menu (6½), Hello (2), What can I do (2).

The above commands were used as possible responses in the multiple choice survey, and one option was added as a special case because this is often used in commercial PC dictation systems: "What can I say". The accompanying question in the second experiment was:

Question 1 "Speak-to-me" is waiting for you to say a command but you do not know which commands are available to you. Which one of the following commands is the most appropriate here?

The Multiple Choice experiment gave rise to the following statistics, tabulated in table A.2

Table A.2: Response data for the Multiple Choice test

Choice	Percentage answered
What can I say	17,6 %
What can I do	5,9 %
Choices	5,9 %
Options	29,4 %
Menu	0,0 %
List commands	35,3 %

Based on the multiple choice statistics, discriminability evidence and predictions, and joint expert opinions, the standardized command was chosen to be "Options" (see table 1).

Annex B (informative): Bibliography

- Baber, C. and Noyes, J.M. (ed.): "Interactive Speech Technology: Human factors issues in the application of speech input/output to computers". Taylor and Francis, 1993.
- DSR/HF-00019 (2001-03, v 0.0.2): "Human Factors (HF): An annotated bibliography of documents dealing with Human Factors and disability".
- Ericsson internal report: "Usage frequency of supplementary services in public and private networks".
- ETSI ETR 170 (1995): "Human Factors (HF): Generic user control procedures for telecommunication terminals and services".
- ETSI ETR 096 (1993): "Human Factors (HF); Phone based interfaces (PBI): Human factors guidelines for the design of minimum phone based user interface to computer services".
- ETSI ETS 300 788 (1997): "Human Factors (HF); Minimum Man-Machine Interface (MMI) to public network based supplementary services".
- Guzman, S.J. et al: "Determining a set of acoustically discriminable, intuitive command words". Proceedings AVIOS, p. 242-250, 2001.
- MacDermid, C. and Goldstein, M: "The "Storyboard" Method: Establishing an Unbiased Vocabulary for Keyword and Voice Command Applications." HCI Industry Day and Adjunct Proceedings, pp 104-109, 1996.
- SpeechDat- Car: "Technical Report LE4-8334-SD1.12: Specification of the car speech database (definition of corpus, scripts and standard), Car environments and speaker coverage", 2001.
- "Speech technology applications for disabled and elderly people": Proceedings of the COST 219 seminar Oberlinhaus, Potsdam-Babelsberg March 21, 1995.
- SPEECON (Speech Driven Interfaces for Consumer Applications) deliverables:
    - "D13: Functionalities of Speech Driven Interfaces";
    - "D21: Specification of Databases- Specification of Language Dependant Items";
    - "D215: Specification of Databases- Specification of Speakers".
- Telephone Speech Standards Committee, Common Dialog Tasks Subcommittee: "Universal Commands for Telephony-Based Spoken Language Systems". SIGCHI Bulletin, Volume 32/2, April 2000.
- ITU-T Recommendation I.210: "Principles of telecommunication services supported by an ISDN and the means to describe them".
- ETSI ETR 095: "Human Factors (HF); Guide for usability evaluations of telecommunications systems and services".
- ETSI ETS 300 738: "Human Factors (HF); Minimum Man-Machine Interface (MMI) to public network based supplementary services".
- ETSI ETR 329: "Human Factors (HF); Guidelines for procedures and announcements in Stored Voice Services (SVS) and Universal Personal Telecommunication (UPT)".
- ITU-T Recommendation F.902: "Interactive services design guidelines".

History

Document history		
V1.1.1	July 2002	Membership Approval Procedure MV 20020920: 2002-07-23 to 2002-09-20
V1.1.1	September 2002	Publication