Methodological considerations for the evaluation of TTS AD's acceptance in the Catalan context

by Anna Fernández-Torné &
Anna Matamala

ARSAD (Advanced Research Seminar) conference 2013.

Barcelona, 13-14/03/13.

Funded by: APOSTA 2011-010, FFI 2012-31024

Methodological considerations for the evaluation of TTS AD's acceptance in the Catalan context

Authors: Anna Fernández Torné & Anna Matamala
Universitat Autònoma de Barcelona, CAIAC

Centre d'Accessibilitat
i Intel·ligència Ambiental
de Catalunya

PhD thesis: the TECNACC project

Technology for Accessibility. Strategies for the Automatisation of the Translation and Voicing of Audio Description

Main goal

In order to

- increase the number of AD products
- improve media accessibility

our **goal** is to **optimise current practices in AD** as regards:

- AD script translation
- AD voice recording

with the help of technologies such as

- MT for the AD script translation
- TTS for the AD voice recording

PhD thesis in 2 stages

TTS AD	April 2012 to April 2013	Evaluation of TTS AD's acceptance in the Catalan context
MT AD	April 2013 to April 2014	Evaluation of MT AD vs. human translated AD

Principles

- **Scientific approach, deductive process:**
  1. Depart from theory
  2. Make hypotheses
  3. Collect data
  4. Obtain findings
  5. Confirm or reject hypotheses
  6. Revise theory
- **Try to reuse of all materials for the different stages of the PhD**

"Social research is often a lot less smooth than the accounts of the research process you read in books like this. [...] In fact, research is full of false starts, blind alleys, mistakes, and enforced changes to research plans." (Bryman, 2012:15)

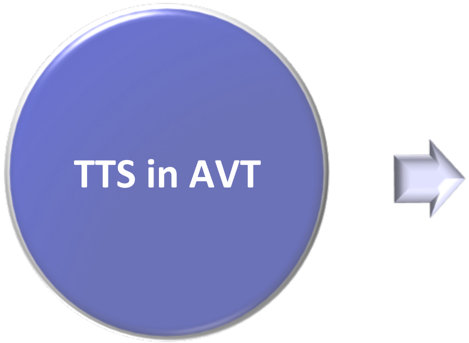
Previous research as theory

TTS in Catalan

- Engines developed by
  - ✓ Loquendo
  - ✓ Nuance
  - ✓ Verbio
  - ✓ iSpeech
  - ✓ eSpeak
  - ✓ Festival

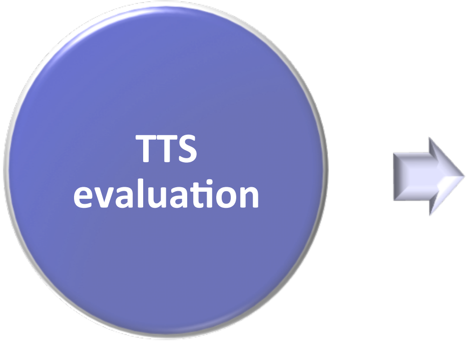
- Alías, Iriondo and Claudi (2011): degree of implementation of TTS in the audiovisual production in Catalonia.

Previous research as theory



- Verboom et al. (2002): Spoken subtitles
- Derbring, Ljunglöf i Olsson (2009): SubTTS
- Orero & Serrano CAIAC: TTS AD + subtitling + spoken subtitles = Universal Accessibility System (UAS)
- Martínez & Mieskes (2011): A Web-based Editor for Audio Titling using Synthetic Speech
- Kobayashi et al. (2010) Are Synthesized Video Descriptions Acceptable?
- Polish TTS AD project:
  - Szarkowska (2011): TTS AD to a monolingual feature film
  - Walczak & Szarkowska (2010): TTS AD to dubbed educational TV series for children
  - Szarkowska & Jankowska's (2012): TTS AD with VO to a foreign fiction film
  - Mączyńska (2011): TTS AD with AST to a documentary
  - TTS AD to a dubbed feature film: in progress

Previous research as theory


TTS evaluation

- International Telecommunitation Union (ITU) Recommendation P.85 (1994): testing method definition for evaluating the subjective quality of synthetic speech. Mean Opinion Score scales.
- Huang, Acero i Hon 2001: taxonomy of evaluation systems.
- Vázquez, Y. i Huckvale, M. (2002) The reliability of the ITU-T P.85 Standard for the evaluation of text-to-speech systems
- Viswanathan & Viswanathan (2004): development and assessment of a modified mean opinion score (MOS) scale
- Sityaev, D., Knill, K. i Burrows, T. (2006) Comparison of the ITU-T P.85 Standard to Other Methods for the Evaluation of TTS systems
- Cryer &Home (2010) review of existing literature "to determine the various methods used for evaluating synthetic voices"
- Cryer, Home & Morley Wilkins (2010): an evaluation protocol
- Hinterletiner et al. (2011) An Evaluation Protocol for the Subjective Assessment of TTS in Audiobook Reading Tasks

Hypothesis

TTS AD will be accepted in the Catalan context

How to confirm the hypothesis?

- Research design: experiment (quasi-experiment)
- Research strategy: quantitative
- Research method: questionnaire

2-level experimental design

1st test

VOICE SELECTION

Subjective quality assessment of synthetic voices: one "best"

Subjective quality assessment of natural voices: one "best"

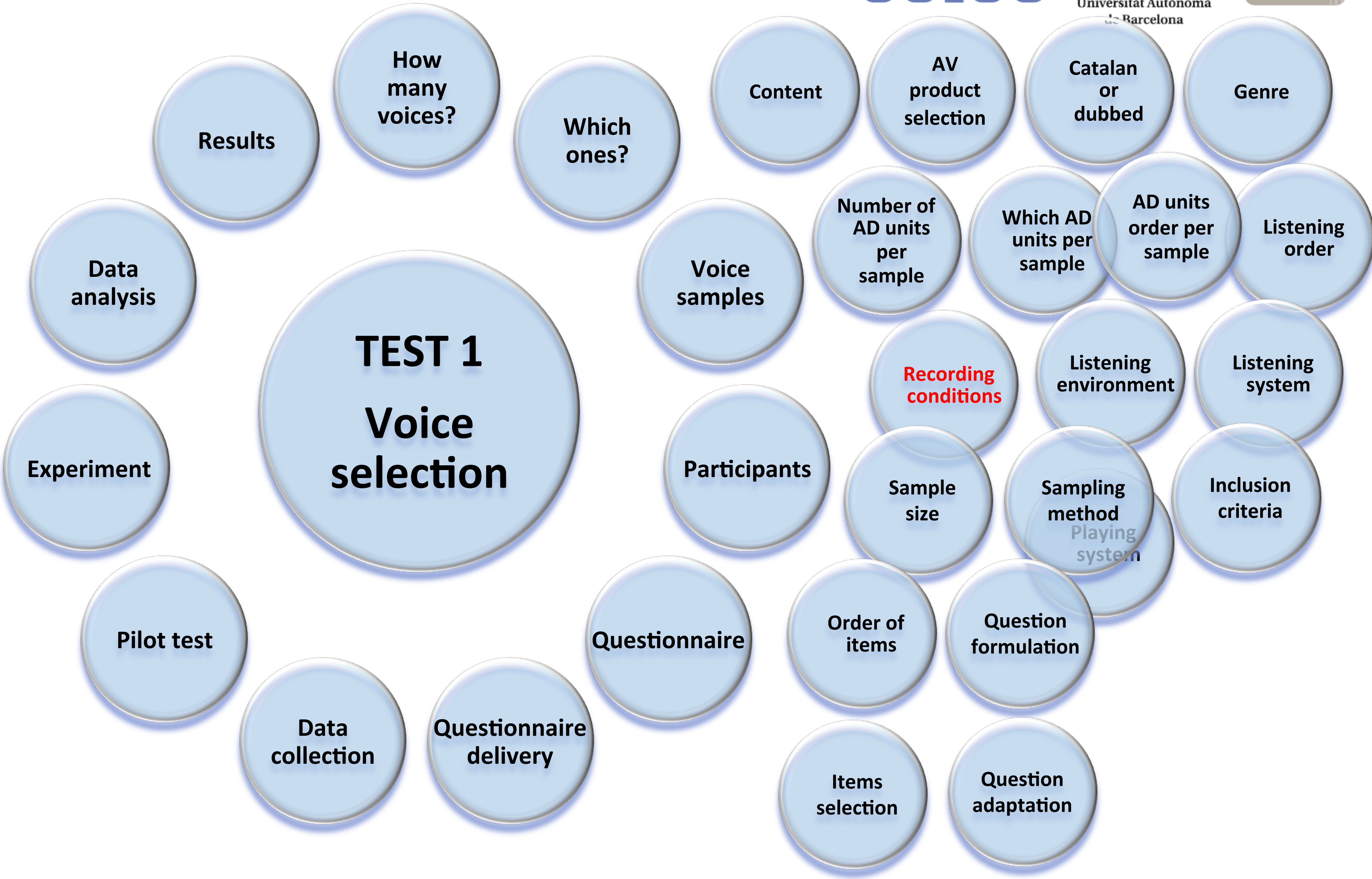
Sighted participants

2nd test

TTS AD ACCEPTANCE

Comparison of subjective quality assessments of an AD product voiced both by the "best" synthetic voice and the "best" natural voice

Blind and partially sighted participants



TEST 1

Voice selection

Results

How many voices?

Which ones?

Voice samples

Content

AV product selection

Catalan or dubbed

Genre

Number of AD units per sample

Which AD units per sample

AD units order per sample

Listening order

Participants

Recording conditions

Listening environment

Listening system

Sample size

Sampling method

Playing system

Inclusion criteria

Questionnaire

Order of items

Question formulation

Items selection

Question adaptation

Questionnaire delivery

Data collection

Pilot test

Experiment

Data analysis

How many?

Voices

- ITU-T P.800 (1996:17) “It is essential for more than one male and more than one female voice to be used in a balanced design.”

5 male / 5 female

- ITU-T P.85 (1994:2) “If possible at least five different sources are recommended, depending on the systems to be tested, applications involved and experimental design.”

x artificial / x natural

20 voices

Which ones?

Synthetic Voices

Voices

Available engines in Catalan:

- Loquendo: Jordi and Montserrat
- Verbio: Oriol and Meritxell
- Acapela: Laia
- iSpeech: Anna
- Nuance: Núria
- FestCat: 5 masculine and 5 feminine voices
- eSpeak: 1 masculine and 1 feminine voices

Selected ones:

- Loquendo: Jordi and Montserrat
- Verbio: Oriol and Meritxell
- Acapela: Laia
- iSpeech: Anna
- FestCat: Pep, Jan, Teo and Ona

Nuance not available

eSpeak very poor quality (almost unintelligible)

Which ones?

Natural Voices

Voices

Disregard of the original AD Catalan voice

Volunteer professional and non professional voice talents

- Generic purposive sampling (not probability sampling), based on individuals who met several criteria (volunteers answering a cold calling email from the Escola Catalana de Doblatge ECAD, either workers, teachers, students, ex students)

- Snowball sampling: the respondents recruited first help localize other participants with similar characteristics

- Inclusion criteria:
    - Having Catalan as mother tongue (central variant)
    - No speech deficiencies

Special thanks to Iola Ledesma and José Carlos Navarro, from ECAD

Voice samples

Content

ITU-T P.85 (1994:1) The messages transmitted by the systems should be related to practical applications.

application related => AD

If an AD is needed, an AV product is also needed

AV product selection

- <u>**To be used both in the TTS AD and MT AD experiments**</u>:
  - Source language EN, with AD in EN
  - Dubbed into CA, with AD in CA
- Just fiction feature films and children's animations are AD in CA
- Corpus chosen

Original title	**Catalan title**
The Bucket List	*O ara o mai*
The Curious Case Of Benjamin Button	*El curiós cas de Benjamin Button*
The Devil Wears Prada	*El diable es vesteix de Prada*
Rocky Balboa	*Rocky Balboa*
Memoirs of a Geisha	*Memòries d'una geisha*
Poseidon	*Posidó*
License to Wed	*Llicència per casar*
Closer	*Closer*

- Final selection: *Closer*, due to availability of all components and miscellaneous genre

Number of AD units

l'ITU-T P.800 (1996:14) “The experimenter must decide how many sentences are required in each group to constitute a speech sample. A minimum of two and a maximum of five are recommended.”

Viswanathan i Viswanathan (2005:65) “Clearly, the larger the number of sentences used the better the results because we find that the synthesis quality is often a function of sentence chosen as it is the system from which it was derived. Synthesis systems are very inconsistent in that certain combinations of phonemes may engender some audio artifacts while other phoneme or word combinations are extremely smooth and natural. Synthesizing multiple sentences is one way to capture some of these idiosyncrasies.”

5 AD units per sample

Voice samples

AD units selection

- l'ITU-T P.800 (1996:14) "Very short and very long sentences should be avoided, the aim being that each sentence when spoken should fit into a time-slot of 2–3 seconds."
- Viswanathan i Viswanathan (2005: 65) "Ideally, synthesized sentences played to listeners to be rated must mimic the application domain in which the TTS system is likely to be used – in sentence length and complexity."
- AD script for *Closer*: 250 AD units, with very different lengths
- Random selection of AD units based on number of characters (no number of words since these might have different amount of characters and cause distortions)

Voice samples

AD units selection

- Viswanathan i Viswanathan (2005: 65) “The same set of sentences is used as input to all of the synthesizers in a study”
- 1 participant would therefore listen to the same set of sentences 10 times in a row => fatigue and learning effect
- Two different set of sentences: one set for the feminine voices and another set for the masculine ones => same content for all feminine/masculine voice samples to be able to compare them properly
- In depth analysis of the AD script and AD units to get to a balance between the two sets of sentences as far as number of characters is concerned.

Voice samples

AD units order

Randomization of AD units within a voice sample (fatigue and practice)

A	1	Fosa a blanc.
	2	Avança fins a l'escenari següent, on tres noies ballen al voltant d'una barra vertical. Mentre una penja cap per avall de la barra, una altra s'ajup amb el cul enfora i un client li posa un bitllet sota la cintura del tanga.
	3	Ella el segueix a l'interior de l'edifici i l'escodrinya amb la mirada.
	4	Se li acosta i li arregla el nus de la corbata.
	5	El noi mig somriu i s'hi encamina mentre ella ho desa somrient al maletí i en treu una poma. Té una ferida al front.
B	2	Avança fins a l'escenari següent, on tres noies ballen al voltant d'una barra vertical. Mentre una penja cap per avall de la barra, una altra s'ajup amb el cul enfora i un client li posa un bitllet sota la cintura del tanga.
	3	Ella el segueix a l'interior de l'edifici i l'escodrinya amb la mirada.
	4	Se li acosta i li arregla el nus de la corbata.
	5	El noi mig somriu i s'hi encamina mentre ella ho desa somrient al maletí i en treu una poma. Té una ferida al front.
	1	Fosa a blanc.
C	5	El noi mig somriu i s'hi encamina mentre ella ho desa somrient al maletí i en treu una poma. Té una ferida al front.
	4	Se li acosta i li arregla el nus de la corbata.
	1	Fosa a blanc.
	2	Avança fins a l'escenari següent, on tres noies ballen al voltant d'una barra vertical. Mentre una penja cap per avall de la barra, una altra s'ajup amb el cul enfora i un client li posa un bitllet sota la cintura del tanga.
	3	Ella el segueix a l'interior de l'edifici i l'escodrinya amb la mirada.
D	4	Se li acosta i li arregla el nus de la corbata.
	1	Fosa a blanc.
	5	El noi mig somriu i s'hi encamina mentre ella ho desa somrient al maletí i en treu una poma. Té una ferida al front.
	3	Ella el segueix a l'interior de l'edifici i l'escodrinya amb la mirada.
	2	Avança fins a l'escenari següent, on tres noies ballen al voltant d'una barra vertical. Mentre una penja cap per avall de la barra, una altra s'ajup amb el cul enfora i un client li posa un bitllet sota la cintura del tanga.
E	3	Ella el segueix a l'interior de l'edifici i l'escodrinya amb la mirada.
	5	El noi mig somriu i s'hi encamina mentre ella ho desa somrient al maletí i en treu una poma. Té una ferida al front.
	2	Avança fins a l'escenari següent, on tres noies ballen al voltant d'una barra vertical. Mentre una penja cap per avall de la barra, una altra s'ajup amb el cul enfora i un client li posa un bitllet sota la cintura del tanga.
	1	Fosa a blanc.
	4	Se li acosta i li arregla el nus de la corbata.

Listening order

Randomization of voice samples to form playlists (order-of-presentation effect), alternating genres

Participant										
1	A	F	B	G	C	H	D	I	E	J
2	B	F	C	G	D	H	E	I	A	J
3	C	G	D	H	E	I	A	J	B	F
4	D	G	E	H	A	I	B	J	C	F
5	E	H	A	I	B	J	C	F	D	G
6	E	H	D	I	C	J	B	F	**A**	G
7	D	I	C	J	B	F	A	G	E	H
8	C	I	B	J	A	F	E	G	D	H
9	B	J	A	F	E	G	D	H	C	I
10	A	J	E	F	D	G	C	H	B	I
11	F	A	G	B	H	C	I	D	J	E
12	G	A	H	B	I	C	J	D	F	E
13	H	B	I	C	J	D	F	E	G	A
14	I	B	J	C	F	D	G	E	H	A
15	J	C	F	D	G	E	H	A	I	B
16	J	C	I	D	H	E	G	A	F	B
17	I	D	H	E	G	A	F	B	J	C
18	H	D	G	E	F	A	J	B	I	C
19	G	E	F	A	J	B	I	C	H	D
20	F	E	J	A	I	B	H	C	G	D

Recording conditions

- Les condicions de gravació han de ser iguals o al més similars possible per a totes les veus (ITU-T P.800 p. 14 In order to eliminate unwanted variability in the speech source, samples of speech having the desired standardized properties should first be prepared in recorded or stored form, as follows.)
- Per a les veus naturals: **Recording environment igual ECAD**
- Per a les veus sintètiques no mateixes qualitats, cosa que es fa palès al resultat final de les mostres sintètiques.

Voice samples

Playing and listening systems

- Playing system: Intel Core 2, 4GB RAM, 180GB RAM, floppy drive, CD-ROM, DVD, USB, network connection and Windows XP. Sound card: Realtek ALC262 @ Intel 82801IB ICH9 - High Definition Audio Controller [A-2] PCI.
- Listening environment: laboratory environment, multimedia classrooms E and D of the Translation and Interpreting Faculty (UAB)
- Listening system: Plantronics Audio 400 DSP

headphones with USB connection.

Participants

Sample size

Huang, Acero i Hon (2001) “Human-subject judgment testing for TTS can adapt methods from speech-coding evaluation (see Chapter 7). With speech coders, *Mean Opinion Score* (MOS) is administered by asking 10 to 30 listeners to rate several sentences of coded speech on a scale of 1 to 5 (1 = Bad, 2 = poor 3 = Fair, 4 = Good, 5 = Excellent)”

Not main experiment:

20 participants

Participants

Sampling method

- Generic purposive sampling (not probability sampling), based on individuals who met several criteria (volunteers answering a cold calling email)
- Snowball sampling: the respondents recruited first help localize other participants with similar characteristics
- ITU-T P.800 (1996:) “No steps are taken to balance the numbers of male and female subjects unless the design of the experiment requires it”

Participants

Inclusion criteria

- Sighted
- Catalan speakers
- No hearing impairments
- No experience in TTS
  - Difficult to control TTS expertise: the more you use TTS applications, the more you understand them and accept them.
- No experience in AD
  - Difficult to control AD expertise

Questionnaire

Items selection (1)

- International Telecommunitation Union Recommendation P.85 (1994): testing method definition for evaluating the subjective quality of synthetic speech (MOS scales) - **8 items**

Type I and Q

- overall impression
- acceptance

Type I

- listening effort
- comprehension problems
- articulation

Type Q

- pronunciation
- speaking rate
- voice pleasantness

Items selection (2)

- Viswanathan & Viswanathan (2005): development and assessment of a modified mean opinion score (MOS) scale – **11 items**

Intelligibility & Naturalness

- overall impression
- acceptance

Intelligibility

- listening effort
- pronunciation
- comprehension
- articulation
- speaking rate

Naturalness

- naturalness
- ease of listening
- pleasantness
- audio flow

Questionnaire

Items selection (3)

- Cryer, Home and Wilkins (2010), four types of assessment (functionality, **subjective, user testing** and technical) in four scenarios (**audio book**, product, document containing figures, access technology application) – **12 items**
  - overall impression, pleasantness, comprehension, pronunciation, prosody, comfortable to listen to for a long period, **responsiveness**, speaking rate, naturalness, listening effort, **appropriate tone**, acceptance

Questionnaire

Items selection (4)

- Hinterleitner et al. (2011), evaluation protocol for the subjective assessment of TTS in audiobook reading tasks – **from 11 to 8 items**

Prosody & Listening pleasure

- overall impression
- comprehension problems
- content
- level of familiarity

Prosody

- intonation
- speech pauses
- emotion
- accentuation

Listening pleasure

- voice pleasantness
- listening effort
- acceptance

Questionnaire

Items selection (5)

Items related to ITU's quality factor rather than intelligibility – **9 items**

- overall impression
- listening effort /ease of listening / comfortable to listen to for a long period
- acceptance
- accentuation
- pronunciation
- speech pauses / audio flow / prosody
- intonation
- naturalness
- voice pleasantness

- **comprehension**
- **speaking rate**

Questionnaire

Order of items

- Randomization of questions is not feasible
- Fix order, from more specific to broader questions

1. Overall impression → opinion
2. Accentuation → words
3. Pronunciation → words
4. Speech Pauses → sentence
5. Intonation → sentence
6. Naturalness → voice
7. Pleasantness → voice
8. Listening effort → opinion
9. Acceptation → opinion

Questionnaire

Question formulation

- Comparison of question formulation in different questionnaires
- Cryer, Home and Wilkins (2010) keep the same structure for the questions and always repeat the same labels for the answers => too tiring and monotonous (9 questions x 10 voices)
- Use of synonym verbs in the questions
- Use of different structures for the questions
- Use of different answer labels for the different questions

Questionnaire

Adaptation into CA

- Careful not to be too technical, clear for everybody
- Appropriate answer labelling: more precise
- Appropriate answer order: like Cryer, Home and Wilkins, since the questionnaire is to be listened to, it is better to go from less positive to more positive (1 to 5), the other way round as ITU and Viswanathan & Viswanathan

Questionnaire delivery

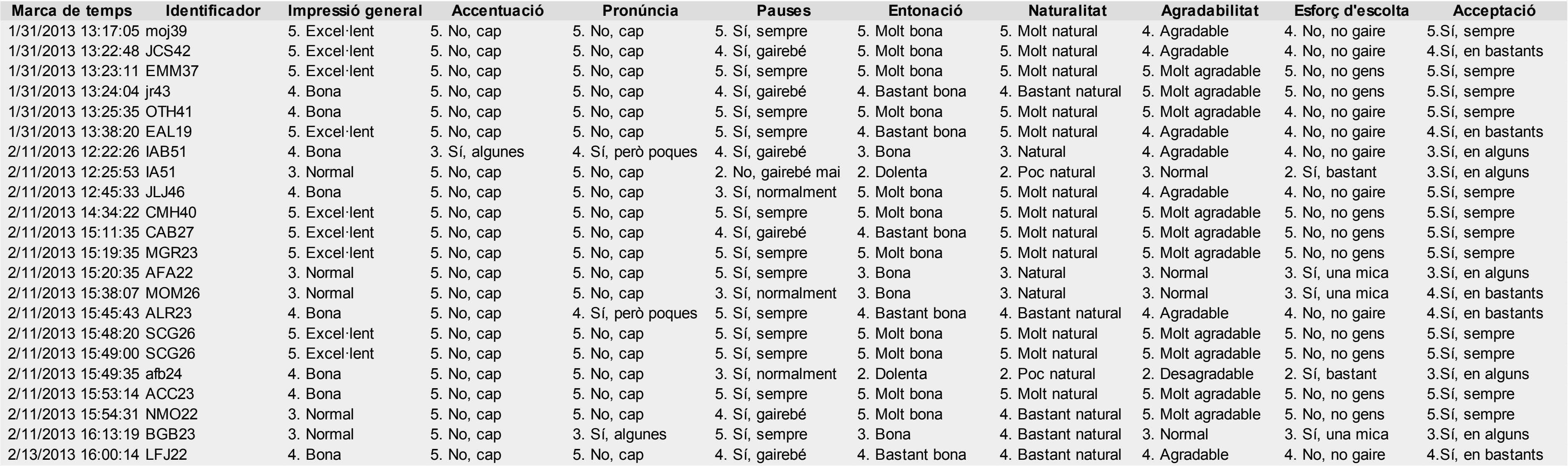
- **Google Forms**: one form per voice
- Questions previously recorded by me
  - 1 second between AD units
  - 5 seconds between voice samples
  - 5 seconds between voice sample and questionnaire
  - 5 seconds between questions: is it enough or too much?
- No inclusion of question titles when reading: too technical (speech pauses), too formal (agradabilitat) or even difficult to be understood (listening effort)

Data collection

Google Forms: one form per voice

i.e. 20 files (one per voice)

including 20 rows each (one row per participant).



Marca de temps	Identificador	Impressió general	Accentuació	Pronúncia	Pauses	Entonació	Naturalitat	Agradabilitat	Esforç d'escolta	Acceptació
1/31/2013 13:17:05	moj39	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	4. Agradable	4. No, no gaire	5.Sí, sempre
1/31/2013 13:22:48	JCS42	5. Excel·lent	5. No, cap	5. No, cap	4. Sí, gairebé	5. Molt bona	5. Molt natural	4. Agradable	4. No, no gaire	4.Sí, en bastants
1/31/2013 13:23:11	EMM37	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
1/31/2013 13:24:04	jr43	4. Bona	5. No, cap	5. No, cap	4. Sí, gairebé	4. Bastant bona	4. Bastant natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
1/31/2013 13:25:35	OTH41	4. Bona	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	4. No, no gaire	5.Sí, sempre
1/31/2013 13:38:20	EAL19	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	4. Bastant bona	5. Molt natural	4. Agradable	4. No, no gaire	4.Sí, en bastants
2/11/2013 12:22:26	IAB51	4. Bona	3. Sí, algunes	4. Sí, però poques	4. Sí, gairebé	3. Bona	3. Natural	4. Agradable	4. No, no gaire	3.Sí, en alguns
2/11/2013 12:25:53	IA51	3. Normal	5. No, cap	5. No, cap	2. No, gairebé mai	2. Dolenta	2. Poc natural	3. Normal	2. Sí, bastant	3.Sí, en alguns
2/11/2013 12:45:33	JLJ46	4. Bona	5. No, cap	5. No, cap	3. Sí, normalment	5. Molt bona	5. Molt natural	4. Agradable	4. No, no gaire	5.Sí, sempre
2/11/2013 14:34:22	CMH40	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:11:35	CAB27	5. Excel·lent	5. No, cap	5. No, cap	4. Sí, gairebé	4. Bastant bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:19:35	MGR23	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:20:35	AFA22	3. Normal	5. No, cap	5. No, cap	5. Sí, sempre	3. Bona	3. Natural	3. Normal	3. Sí, una mica	3.Sí, en alguns
2/11/2013 15:38:07	MOM26	3. Normal	5. No, cap	5. No, cap	3. Sí, normalment	3. Bona	3. Natural	3. Normal	3. Sí, una mica	4.Sí, en bastants
2/11/2013 15:45:43	ALR23	4. Bona	5. No, cap	4. Sí, però poques	5. Sí, sempre	4. Bastant bona	4. Bastant natural	4. Agradable	4. No, no gaire	4.Sí, en bastants
2/11/2013 15:48:20	SCG26	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:49:00	SCG26	5. Excel·lent	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:49:35	afb24	4. Bona	5. No, cap	5. No, cap	3. Sí, normalment	2. Dolenta	2. Poc natural	2. Desagradable	2. Sí, bastant	3.Sí, en alguns
2/11/2013 15:53:14	ACC23	4. Bona	5. No, cap	5. No, cap	5. Sí, sempre	5. Molt bona	5. Molt natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 15:54:31	NMO22	3. Normal	5. No, cap	5. No, cap	4. Sí, gairebé	5. Molt bona	4. Bastant natural	5. Molt agradable	5. No, no gens	5.Sí, sempre
2/11/2013 16:13:19	BGB23	3. Normal	5. No, cap	3. Sí, algunes	5. Sí, sempre	3. Bona	4. Bastant natural	3. Normal	3. Sí, una mica	3.Sí, en alguns
2/13/2013 16:00:14	LFJ22	4. Bona	5. No, cap	5. No, cap	4. Sí, gairebé	4. Bastant bona	4. Bastant natural	4. Agradable	4. No, no gaire	4.Sí, en bastants

Pilot test

- 6 participants, different professions, different degrees of education

  →3 women: 1x20 to 40; 1x40 to 60; 1 more than 60

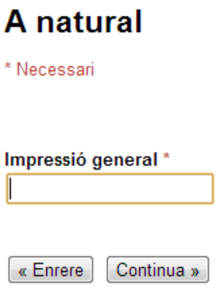
  →3 men: 1x20 to 40; 1x40 to 60; 1 more than 60

- The aim was to validate the design itself, the questions and their CA adaptation, not so much technical specifications
- Laboratory environment, but not the one used in the experiment
- Different PC and headphones as the ones to be used in the experiment
- 1 session: either synthetic or natural voices
  - 3 participants (2 women, 1 man; one from each age range) heard synthetic voices
  - 3 participants (1 woman, 2 men; one from each age range) heard natural voices

Pilot test results

- Contextualisation + recorded instructions: disorganised and repeated info
- Questionnaire with recorded questions: too slow in the end (learning effect)
- Not all participants have adequately understood the questions

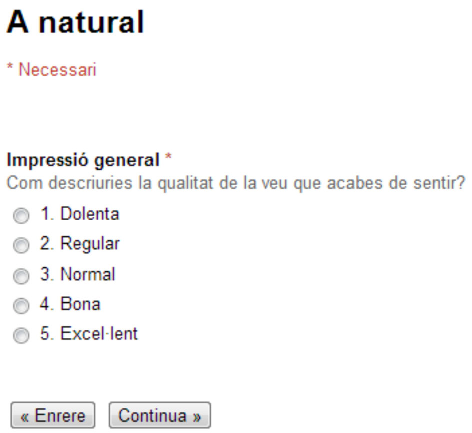
1. Need to add questions and answers to be seen by participants, not just heard => redesigning forms and therefore instructions (the way participants were told to fill in the answers)



A natural

* Necessari

Impressió general *



A natural

* Necessari

Impressió general *
Com descriuries la qualitat de la veu que acabes de sentir?

○ 1. Dolenta
○ 2. Regular
○ 3. Normal
○ 4. Bona
○ 5. Excel·lent

2. Reformulation of questions (not feasible since they should be rerecorded)
3. Addition of detailed explanation of what is intended by each question => rewriting of the contextualisation document including the instructions

Experiment

- 20 participants, as heterogeneous as possible from the profession and degree of education point of view
  - 6 men and 14 women ranging from 19 to 51
- Two sessions: 1. synthetic voices 2. natural voices (otherwise too long)
  - Two sets of sessions needed due to lack of participants

Data analysis

- Analysis software: SAS, v9.2, SAS Institute Inc., NC, USA.

1. Descriptive statistics (mean, median, standard deviation, minimum, maximum, lower quartile and upper quartile) of the scores of the following 8 questions: accentuation, acceptance, pleasantness, intonation, listening effort, naturalness, speech pauses and pronunciation.
2. Overall impression analysis:
    - Descriptive statistics (mean, median, standard deviation, minimum, maximum, lower quartile and upper quartile) of the scores of the overall impression.
    - Overall impression modelling: multinomial model, where the dependent variable is the score of the overall impression and the independent variable is the voice. A **statistical hypothesis test** is carried out to answer the question of whether there are statistically significant differences among voices (H0: There are no differences among voices / H1: There are differences among voices). Once it has been confirmed that there are differences among voices (refuse H0), a 2 to 2 comparison with corrections is made to determine between which voices differences exist, taking into account that each participant has listened to 5 voices for each case (5 feminine natural, 5 feminine synthetic, 5 masculine natural and 5 masculine synthetic).
3. If a best voice cannot be concluded from the overall impression modelling, then the acceptance modelling is performed, where the dependent variable is the score of the acceptance and the independent variable is the voice, taking into account that each participant has listened to 5 voices for each case (5 feminine natural, 5 feminine synthetic, 5 masculine natural and 5 masculine synthetic).

Results

- Pilot test and experiment results match

1. Best feminine natural voice: **D** (**professional voice talent**)
2. Best masculine natural voice: no statistically significant differences among F, H and I, but **F** has a slightly higher score acceptance (not even in overall impression) – **2nd year student voice talent**
3. Best feminine synthetic voice: no statistically significant differences between A and C, but **A** has a higher score in overall impression and acceptance
4. Best masculine synthetic voice: **H**

Thank you!

Ana.Fernandez.Torne@uab.cat
Universitat Autònoma de Barcelona, CAIAC

Centre d'Accessibilitat
i Intel·ligència Ambiental
de Catalunya