

# Nautilus Speaker Characterization (NSC) Corpus

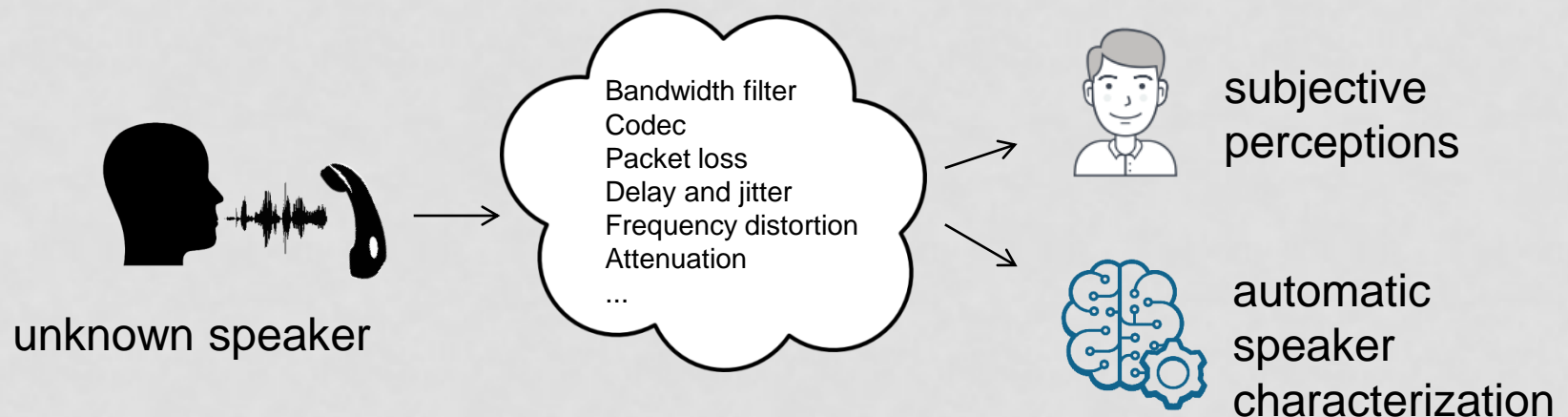
*Laura Fernández Gallardo, Ph.D.*

*Quality and Usability Lab, Technische Universität Berlin, Germany*



# Motivation

- Needed: speech database for speaker characterization
- Social / interpersonal characteristics:
  - e.g. likable, attractive, friendly, competent, modest, childish,...
- Effects of transmission channels
  - Needed: clean microphone speech and at least  $f_s = 32$  kHz



# NSC Corpus overview

Purpose	Speaker characterization
Number of speakers	300 speakers; 126 males, 174 females
Speakers' age	18 to 35 years old
Speakers' language	German (mother tongue)
Speech type	conversational; scripted, semi-spontaneous, and spontaneous
Speech duration; size	155 hours of speech; 50 GB
File format	1-channel audio/wav, 48 kHz, 16 bit speakers: microphone AKG C 414B-XLS interlocutor: headset Sennheiser HMD 46
Metadata	demographic information and self-assessed personality
Labels	34 interpersonal speaker attributions (continuous numeric) 34 voice descriptions for selected speakers (continuous numeric)
ISLRN	<a href="https://www.islrn.org/record/157-037-166-491-1">157-037-166-491-1</a>
Author / data owner	<a href="#">Laura Fernández Gallardo</a>

# Speech Material

- 4 scripted dialogs
  - Dialog 1: Health insurance inquiry
  - Dialog 2: Mobile phone rate plan inquiry
  - Dialog 3: Car rental inquiry
  - Dialog 4: Real estate agency inquiry
- 4 semi-spontaneous dialogs
  - Dialog 5: Car rental booking
  - Dialog 6: Pizza order
  - Dialog 7: Book from the library
  - Dialog 8: Doctor's appointment
- Spontaneous speech: casual neutral, emotional, question and statement utterances
- Interlocutor: student worker
  - speaker *amsterdam* in 279 sessions
  - speaker *aden* in 21 sessions

# Speech Material

## scripted dialogs

Dialog 1: Krankenversicherung		Guten Tag, DKS Versicherungen. Mein Name ist Heinmüller, was kann ich für Sie tun?
	1a	<b>Guten Tag, mein Name ist Schmidt. Ich hab eine Frage bezüglich meiner Krankenversicherung.</b> Geben Sie mir bitte Ihre Vertragsnummer durch.
	1b	<b>Meine Vertragsnummer ist die 4035.</b> Einen Moment bitte. Alles klar, wie lautet Ihre Frage?
	1c	<b>Ich hab von meinem Hausarzt ein Rezept über 10 Physiotherapie-Anwendungen bekommen. Deshalb wollte ich mal fragen, ob Sie die Kosten in vollem Umfang übernehmen.</b> Ich verstehe. Ich leite Sie eben zu Ihrem persönlichen Sachbearbeiter weiter.
	1d	<b>Alles klar, Dankeschön!</b>

## semi-spontaneous dialogs

### Autovermietung

Ihr Name: *Pfeifer*



1 Kleintransporter für einen Umzug



für ein Wochenende (Samstag und Sonntag),  
mindestens 1.80 m Ladehöhe,  
möglichst günstig



Modell : \_\_\_\_\_

Ladehöhe : \_\_\_\_\_

Preis : \_\_\_\_\_ €/Tag



Reservierung : Eurocard

Kreditkartennummer: 8823 7453 6592 0023

gültig bis 12/19

Führerschein Klasse 3

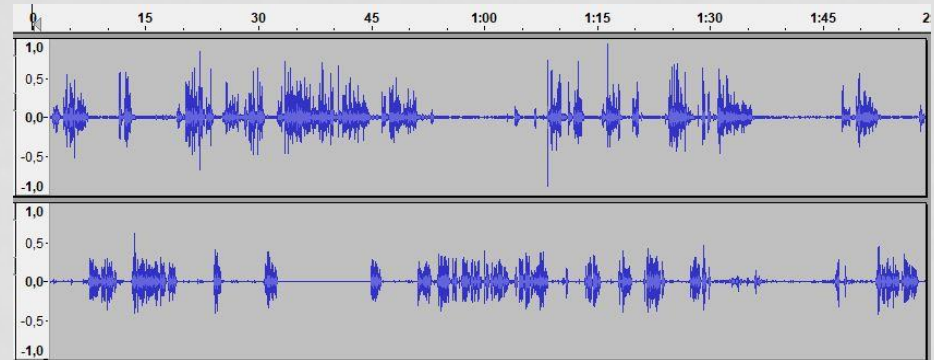


Ab wieviel Uhr kann ich den Wagen morgens abholen?

\_\_\_\_\_

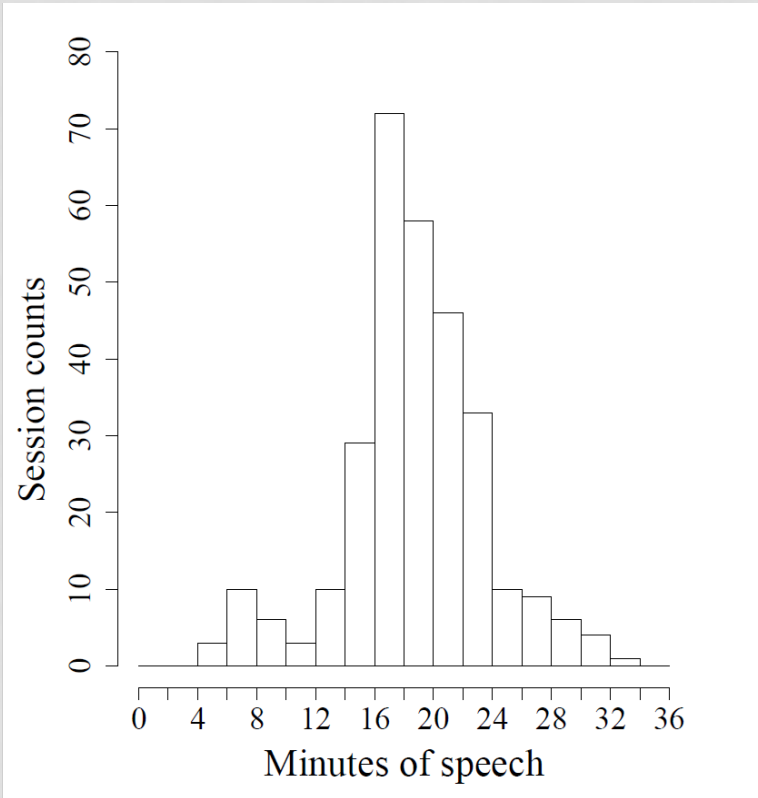
# Speech Material

- Interactions:
  - interlocutor speech (wav)
  - speaker speech (wav)
  - tags+timestamps (csv)
    - d: dialog turn
    - s: neutral spontaneous speech
    - e: spontaneous emotional speech
    - q: spontaneous questions
    - f: spontaneous short feedback
- Complete interactions from the start of Dialog 5 until the end of Dialog 8
- Other short interactions when spontaneous speech



#	tag	timestart_s	timeend_s
1	5	7.329	11.676
2	5	13.058	19.602
3	s	23.980	25.606
4	5	30.780	33.060
5	5	44.737	46.808
6	5	51.009	53.433
7	5	53.679	64.028
8	s	64.028	68.978
9	5	73.159	75.942
10	5	77.635	80.046
11	5	81.325	84.458
12	5	87.137	90.534
13	f	106.631	108.604
14	6	111.893	117.951

# Speech Material



Distribution of the session's speech  
(not including interlocutor's speech or stimuli files)

- Duration of all speakers' speech:
  - 18.2 h of scripted speech
  - 14.5 h of (sorted) semi-spontaneous speech
  - 59.5 h of interactions
  - 1.2 h of concatenations
  - 0.75 h of sustained vowel 'a'
  - 1.9 h of speech stimuli employed for labeling
- Duration of interlocutor's speech:
  - 54.2 h for *amsterdam*
  - 5.3 h for *aden*
- Thus, total  $\approx$  155 h of speech

# Speakers' metadata

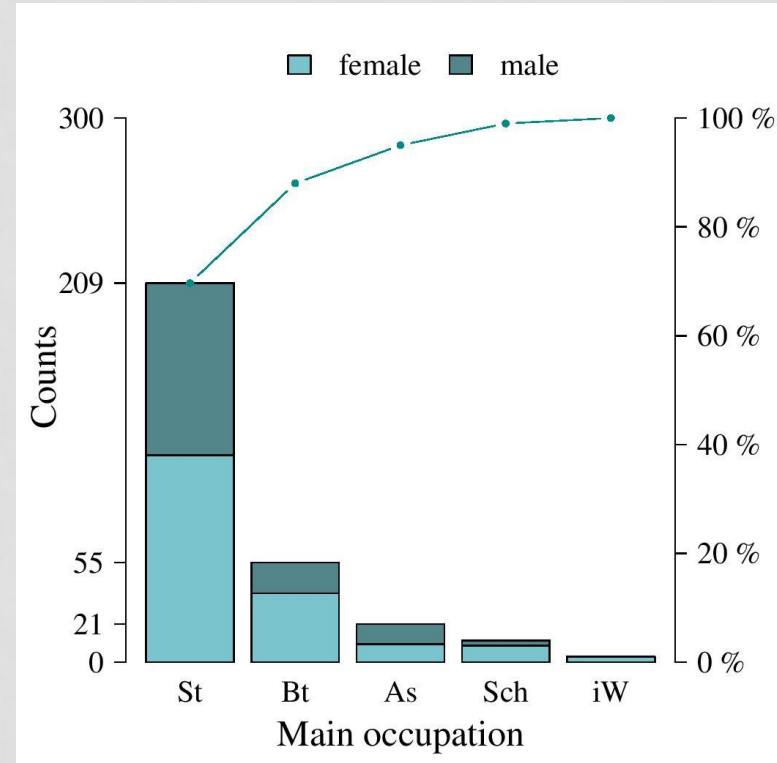
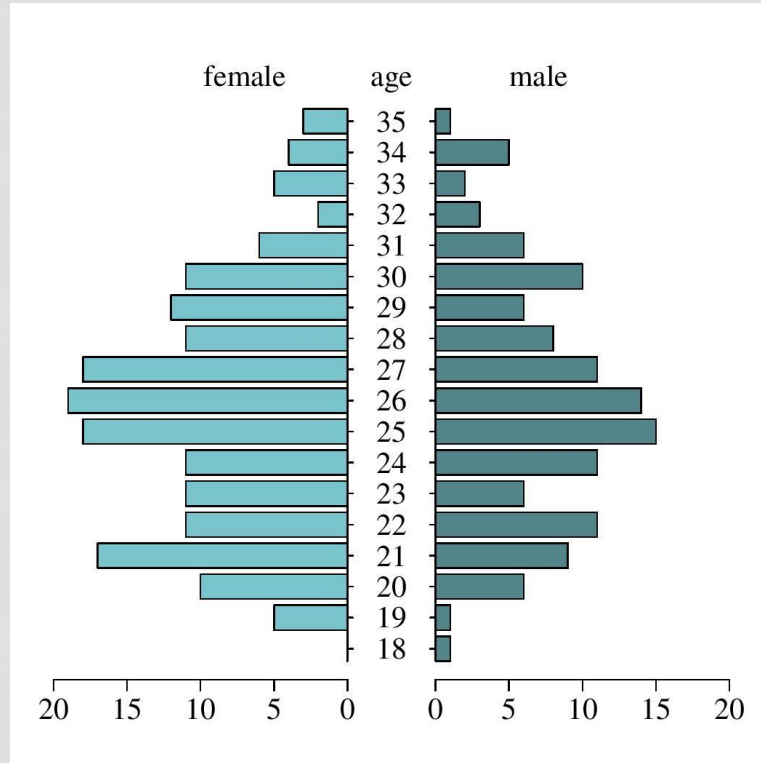
- Speakers' recruitment
  - “Probandenportal” of the Technische Universität Berlin
  - mailing lists
  - announcements
- Requirements:
  - Native German speaker without dialect or accent
  - 18 – 35 years old
  - No hearing or speech disorders
  - No sickness or cold at the time of the recording
- Speakers were paid for their participation



# Speakers' metadata

- Socio-demographic data:
  - age
  - gender
  - place of birth
  - chronological places of residence and duration of stay
  - place of birth of the mother and of the father
  - highest education level
  - educational background
  - main occupation
  - past occupation(s) if any
  - years of work experience
- self-assessed personality traits and vocal attractiveness:
  - 92 speakers completed the BFI-10
  - 181 speakers completed the BFI-44
  - 237 speakers answered to: “Wie sympathisch könnten andere Ihre Stimme finden?” (“How likable could others find your voice?”)


# Speakers' metadata



Main occupation: Student, Berufstätig (employed), Arbeitssuchend (seeking employment), Schüler / Auszubildender (trainee), and in Weiterbildung (continuing education)

# Speakers' metadata

- Speakers sign a consent form:
  - All data are pseudo-anonymized
  - The database may be publicly released under license agreement for exclusive use in scientific studies
  - The participation is voluntary and it is possible to withdraw from the study
- Every speaker was assigned a unique city name as pseudonym

**Technische Universität Berlin** 

Einwilligung des Teilnehmers

Ich erkläre, dass

- ich die Informationen zur Studie „Sprachaufnahmen zur Unterstützung sprachwissenschaftlicher Forschung“ gelesen habe und bereit bin, an der Studie teilzunehmen.
- meine Fragen ausreichend und verständlich beantwortet wurden.
- ich darüber informiert wurde, dass von mir heute folgende Daten erhoben werden: Audioaufnahmen und soziodemographische Daten.
- ich darüber informiert wurde, dass sämtliche Daten gespeichert und pseudonymisiert für wissenschaftliche Auswertungen genutzt werden.
- ich mir bewusst bin, dass die Audioaufnahmen und die Metadaten Teil einer öffentlichen Datenbank werden, die auch anderen Wissenschaftlern zu Forschungszwecken zur Verfügung gestellt werden soll. Alle Daten, die von mir für diese Studie erfasst werden, werden vertraulich behandelt.
- Personenbezogene Daten werden vor dem Zugriff Unbefugter geschützt.
- ich mir bewusst bin, dass die Teilnahme freiwillig ist und ich sie jederzeit ohne Angaben von Gründen widerrufen kann, ohne dass mir daraus irgendein Nachteil entsteht.
- ich eine Kopie der Informationen zur Studie und dieser Einwilligung erhalten habe.

Zudem ist mir bekannt, dass alles, was ich während der heutigen Sitzung gesehen und gehört habe, vertraulich ist und erkläre, dass ich diese Informationen nicht weitergeben werde.

Name: \_\_\_\_\_

Datum: \_\_\_\_\_

Unterschrift: \_\_\_\_\_

Versuchsleiterin:  
Laura Fernández Gallardo, Ph.D.  
Quality and Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin  
Ernst-Reuter-Platz 7, D-10587 Berlin, Germany  
+49 175 2629100  
Laura.Fernandez-Gallardo@telekom.de

# Recording Procedure

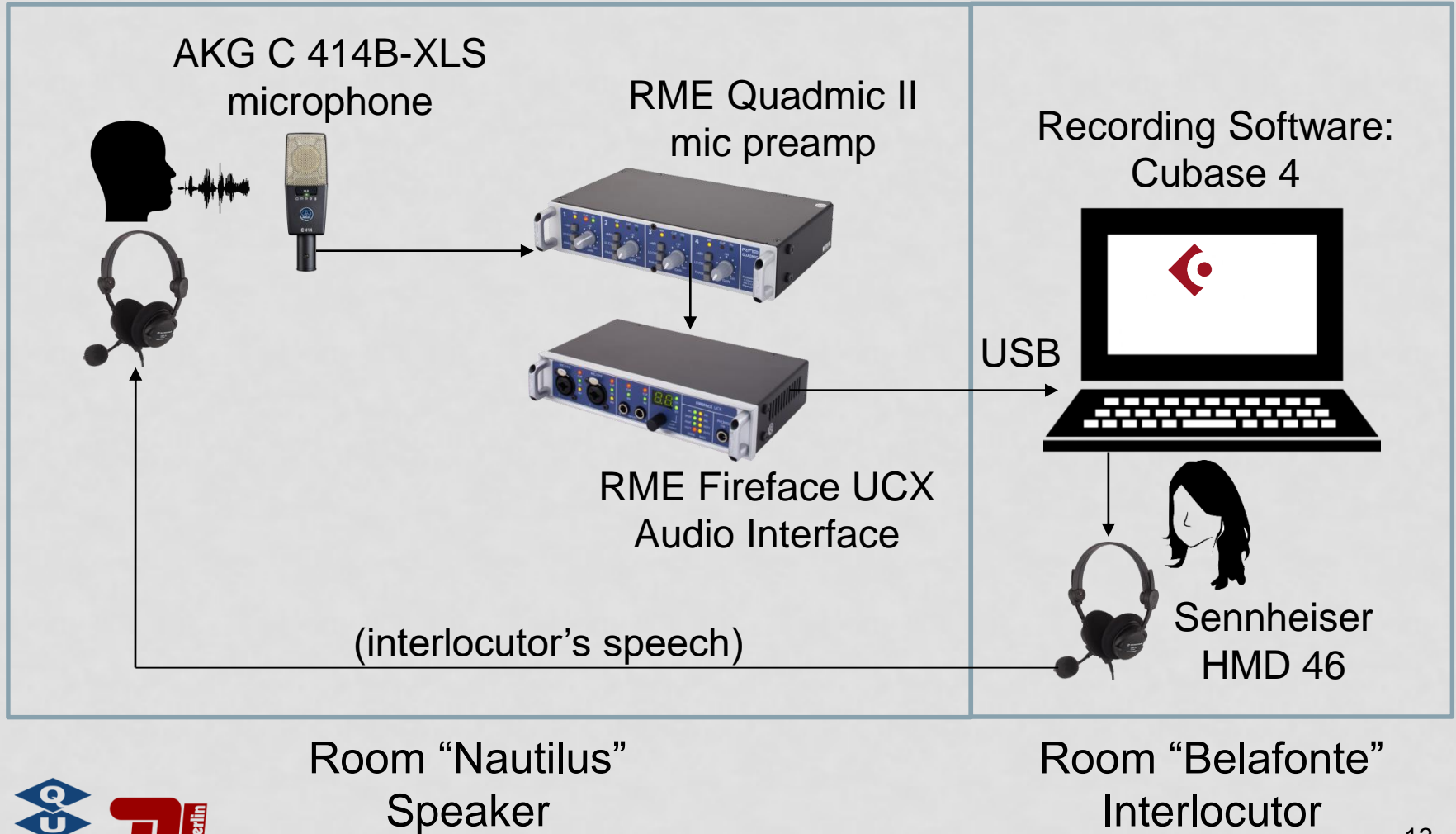
- Speaker sits in Room “Nautilus”
  - acoustically-isolated
  - $RT60 = 0.08$  s at 2 kHz
  - (height x width x depth): 2,10m x 2,75m x 2,53m



(In the position of a speaker is Laura Fernández Gallardo).

- Interlocutor sits in Room “Belafonte”
  - Subject to background noises

# Recording Procedure



# Recording Procedure

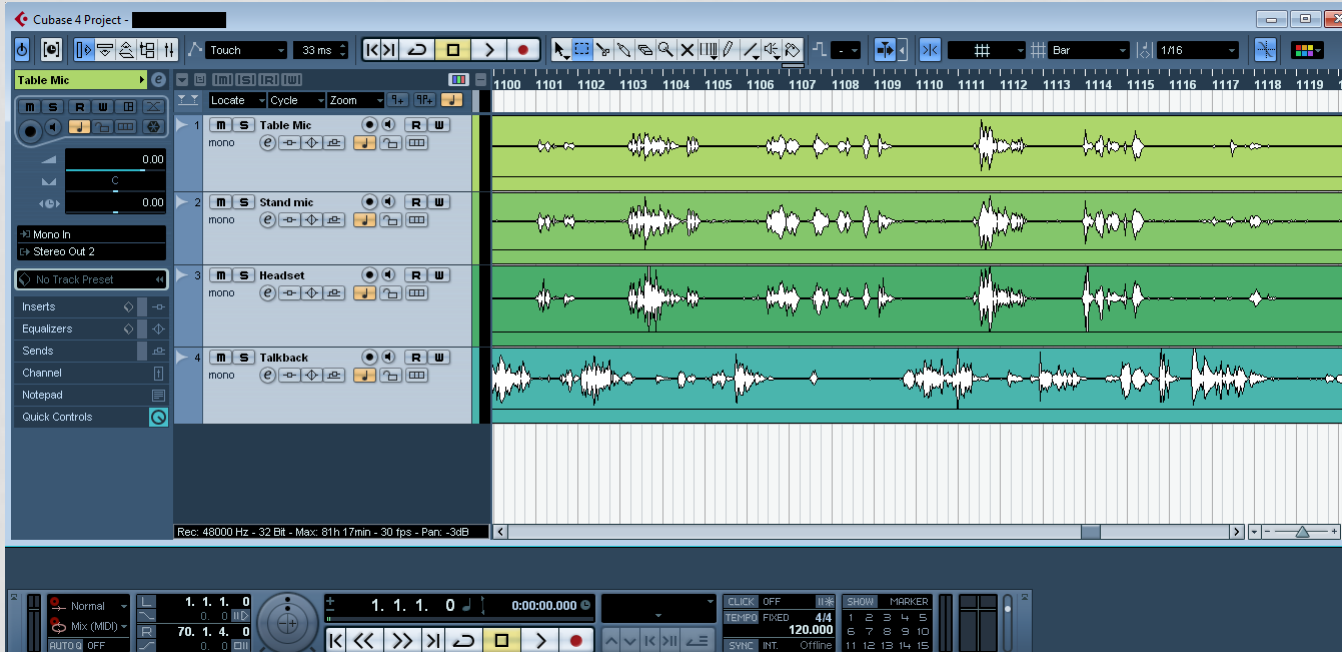


Table microphone:  
Audio-Technica U851R

Stand-up microphone:  
AKG C 414B-XLS

Headset microphone:  
Sennheiser HMD 46

Interlocutor's microphone:  
Sennheiser HMD 46

While other two microphones were used to acquire the speakers' speech at 48 kHz, only the recordings by the AKG mic are included in the *share* version of NSC

# Speech labeling

## Interpersonal Speaker Characteristics SC-Questionnaire

- All 300 speakers evaluated
- 34-item semantic differential rating scale
- Stimuli: Dialog 6 (semi-spontaneous, pizza order)
- Listening tests with 114 participants (70 m, 44 f)
- On average, the SC-Questionnaire was completed 15 times for each speaker by different listeners

What characteristics  
are attributed to the  
speaker?



item	Antonyms (German)		English translation	
1	sympathisch	unsympathisch	likable	non-likable
2	unsicher	sicher	insecure	secure
3	unattraktiv	attraktiv	unattractive	attractive
4	verständnisvoll	verständnislos	sympathetic	unsympathetic
5	entschieden	unentschieden	decided	indecisive
6	aufdringlich	unaufdringlich	obtrusive	unobtrusive
7	nah	distanziert	close	distant
8	interessiert	gelangweilt	interested	bored
9	emotionslos	emotional	unemotional	emotional
10	genervt	nicht genervt	irritated	not irritated
11	passiv	aktiv	passive	active
12	unangenehm	angenehm	unpleasant	pleasant
13	charaktervoll	charakterlos	characterful	characterless
14	reserviert	gesellig	reserved	sociable
15	nervös	entspannt	nervous	relaxed
16	distanziert	mitfühlend	distant	affectionate
17	unterwürfig	dominant	conformable	dominant
18	affektiert	unaffektiert	affected	unaffected
19	gefühlskalt	herzlich	cold	hearty
20	jung	alt	young	old
21	sachlich	unsachlich	factual	not factual
22	aufgeregt	ruhig	excited	calm
23	kompetent	inkompetent	competent	incompetent
24	schön	hässlich	beautiful	ugly
25	unfreundlich	freundlich	unfriendly	friendly
26	weiblich	männlich	feminine	masculine
27	provokativ	gehorsam	offensive	submissive
28	engagiert	gleichgültig	committed	indifferent
29	langweilig	interessant	boring	interesting
30	folgsam	zynisch	compliant	cynical
31	unaufgesetzt	aufgesetzt	genuine	artificial
32	dumm	intelligent	stupid	intelligent
33	erwachsen	kindlich	adult	childish
34	frech	bescheiden	bold	modest

# Speech labeling

Factor analysis for male and for female speakers separately

	Warmth	Attract.	Confid.	Compl.	Matur.
herzlich	0.85				
mitfuehlend	0.84				
distanziert	-0.76				
freundlich	0.59				
verstaendnislos	-0.58				
unsympatisch	-0.52				
nicht.genervt	0.51				
attraktiv		0.85			
haesslich		-0.79			
angenehm		0.58			
interessant		0.48			
sicher			1.00		
unentschieden			-0.6		
gehorsam				0.87	
zynisch				-0.71	
alt					0.82
kindlich					-0.73



	Warmth	Attract.	Compl.	Confind.	Matur.
herzlich	0.84				
mitfuehlend	0.84				
distanziert	-0.78				
freundlich	0.56				
verstaendnislos	-0.49				
nicht.genervt	0.49				
unsympatisch	-0.45				
attraktiv		0.83			
haesslich		-0.81			
angenehm		0.59			
gehorsam			0.8		
zynisch			-0.72		
sicher				0.82	
unentschieden				-0.81	
kindlich					-0.81
alt					0.68



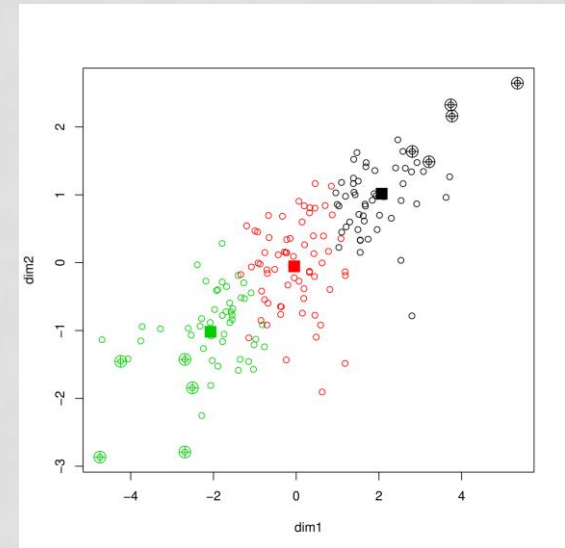
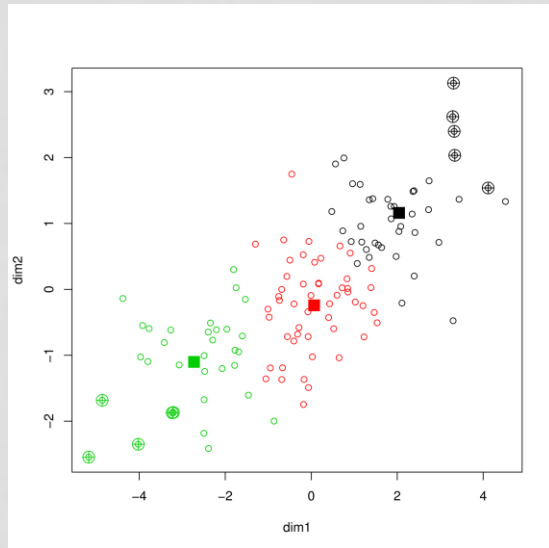
What characteristics are attributed to the speaker?





# Speech labeling

- Selecting 20 “extreme” speakers of the warmth-attractiveness space



What characteristics are attributed to the speaker?



# Speech labeling

## Naïve Voice Descriptions VD-Questionnaire

- For selected 20 “extreme” speakers (10 m, 10 f low/high) in the warmth-attractiveness dimensions
- 34-item semantic differential rating scale
- Same stimuli as for SC: Dialog 6
- Listening tests with 26 participants (13 m, 13 f)
- The VD-Questionnaire was completed by all listeners for the 20 speakers



What are the peculiarities of the speaker's voice?

item	Antonyms (German)		English translation	
1	klangvoll	klanglos	sonorous	flat
2	tief	hoch	low	high
3	nasal	nicht nasal	nasal	not nasal
4	stumpf	scharf	blunt	sharp
5	gleichmäßig	ungleichmäßig	even	uneven
6	akzentfrei	mit Akzent	accented	without accent
7	dunkel	hell	dark	bright
8	leise	laut	quiet	loud
9	knarrend	nicht knarrend	creaky	not creaky
10	variabel	monoton	variable	monotonous
11	angenehm	unangenehm	pleasant	unpleasant
12	deutlich	undeutlich	articulate	inarticulate
13	rau	glatt	coarse	not coarse
14	klar	heiser	clear	hoarse
15	unauffällig	auffällig	not remarkable	remarkable
16	schnell	langsam	quick	slow
17	kalt	warm	cold	warm
18	unnatürlich	natürlich	unnatural	natural
19	stabil	zittrig	stable	shaky
20	unpräzise	präzise	imprecise	precise
21	brüchig	fest	brittle	firm
22	unmelodisch	melodisch	not melodious	melodious
23	angespannt	entspannt	tense	relaxed
24	holprig	gleitend	bumpy	smooth
25	lang	kurz	long	short
26	locker	gepresst	lax	pressed
27	kraftvoll	kraftlos	powerful	powerless
28	flüssig	stockend	fluent	halting
29	weich	hart	soft	hard
30	professionell	unprofessionell	professional	unprofessional
31	betont	unbetont	emphasized	not emphasized
32	sanft	schrill	gentle	shrill
33	getrennt	verbunden	disjointed	jointed
34	nicht behaucht	behaucht	not breathy	breathy

# Folder Structure

```
NSC Corpus - root
├── doc
│   ├── NSC_Documentation.vXX.pdf
│   ├── NSC_License.CLARIN.ACA.BY.NC.NORED.pdf
│   ├── SC_dimensionloadings.csv
│   ├── SC_factorscores_femalespk.csv
│   ├── SC_factorscores_malespk.csv
│   ├── SC_ratings.csv
│   ├── speakers_metadata.csv
│   ├── VD_dimensionloadings.csv
│   ├── VD_factorscores_femalespk.csv
│   ├── VD_factorscores_malespk.csv
│   └── VD_ratings.csv
├── sessions
│   ├── [g] [id] -[pseudonym]
│   │   └── interactions
│   │       ├── [g] [id] -[pseudonym]_interaction_[ii]_interlocutor.wav
│   │       ├── [g] [id] -[pseudonym]_interaction_[ii]_speaker.wav
│   │       ├── [g] [id] -[pseudonym]_interaction_[ii]_speaker_timestamp.wav
│   │       ├── [g] [id] -[pseudonym]_semispontaneousdialogs_interlocutor.wav
│   │       ├── [g] [id] -[pseudonym]_semispontaneousdialogs_speaker.wav
│   │       └── [g] [id] -[pseudonym]_semispontaneousdialogs_speaker_timestamp.wav
│   ├── scripted_turns
│   │   ├── [g] [id] -[pseudonym] -[dd] -[tt] -[nn].wav
│   │   └── semispontaneous_turns
│   │       ├── [g] [id] -[pseudonym] -[dd].wav
│   │       └── sustained
│   │           └── [g] [id] -[pseudonym]_sustained.wav
└── stimuli
    ├── [g] [id] -[pseudonym]_stimulus.wav
```

# Sharing the NSC Corpus

- NSC Corpus ISLRN: [157-037-166-491-1](#)
- Freely released to the scientific community under a license agreement (**CLARIN ACA+BY+NC+NORED**)
- Repositories:
  - [CLARIN](#)
  - [ELRA](#)
  - LDC (coming soon)
- Audio samples can be found [here](#)



# Sharing the NSC Corpus

- By using the corpus resources you agree:
  - to use the NSC corpus **for non-commercial research and teaching purposes only**
  - not to redistribute the NSC corpus or parts of it to third parties
  - to cite the following source in any published work which is based on the corpus:

Fernández Gallardo, L. and Weiss, B., "The Nautilus Speaker Characterization Corpus: Speech Recordings and Labels of Speaker Characteristics and Voice Descriptions," in International Conference on Language Resources and Evaluation (LREC), 2018.

# Supported by

- Supported by DFG, Grant FE 1603/1-1 to Laura Fernández Gallardo
- Cooperating with this study:
  - Dr. Benjamin Weiss (benjamin.weiss@tu-berlin.de)
  - Prof. Dr.-Ing. Sebastian Möller (sebastian.moeller@tu-berlin.de)
- Acknowledgements:
  - *amsterdam, aden*, Benjamin Weiss, Christian Krüger, Christoph Hold, Lars-Erik Riechert, Rafael Zequeira Jiménez

# Thank you for your attention!

*Laura Fernández Gallardo*  
*[laura.fernandezgallardo@tu-berlin.de](mailto:laura.fernandezgallardo@tu-berlin.de)*  
*<http://www.qu.tu-berlin.de/?id=lfernandez>*

**We are open to any form of collaboration.  
Feel free to contact us!**