

Michael Pucher (FTW), Friedrich Neubarth (OFAI), Volker Strom (CSTR),
Sylvia Moosmüller (ARI), Gregor Hofer (CSTR), Christian Kranzler
(FTW), Gudrun Schuchmann (FTW), Dietmar Schabus (FTW)

Telecommunications Research Center Vienna (FTW)

The Austrian Research Institute for Artificial Intelligence (OFAI)

Acoustic Research Institute, Austrian Academy of Sciences (ARI)

Centre for Speech Technology Research, University of Edinburgh (CSTR)

RESOURCES FOR SPEECH SYNTHESIS OF VIENNESE VARIETIES

Contents

- Project „Viennese Sociolect and Dialect Synthesis (VSDS)“
- Viennese varieties
- Synthesis samples
- Voice development
 - Speaker selection
 - Recording
 - Text selection
 - Phone sets
 - Spoken dialog system
- Release 1.0

Project „Viennese Sociolect and Dialect Synthesis“



- **Development of synthetic dialect voices**

- Nationally funded project
- Development of 1 Austrian German and 3 Viennese sociolect voices
- Lexicon development
- Efficient methods for less resourced varieties
- Automatic generation of in-between varieties

- **Scenarios**

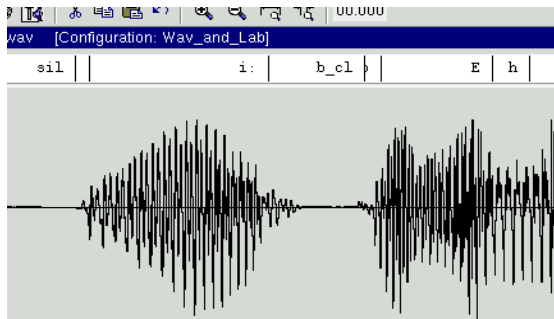
- Scenario research on regionalized services
- Potential applications: tourism, education, gaming
- Location based application: Regionalized restaurant guide for Vienna, where different dialects are associated with different regions/types of restaurants

- **Project partners**

- Telecommunications Research Center Vienna (FTW)
- The Austrian Research Institute for Artificial Intelligence (OFAI)
- Acoustic Research Institute, Austrian Academy of Sciences (ARI)
- Centre for Speech Technology Research, University of Edinburgh (CSTR)

Project homepage:

<http://dialect-tts.ftw.at>



Viennese varieties

- Historically influenced by many languages (Czech, French, Jiddisch,...) as can be seen by the lexicon of Viennese words
- „Viennese dialect“ refers to a sociolect (*education, age, gender*) spoken within a dialectal region
 - previous studies showed that *age* and *educational* level define Viennese sociolects
- Therefore we decided to realize 3 sociolect personas / voices that represent a 3-dimensional sociolect space (*age, gender, education*)





Code	Variety	Speaker	Education	Age group	Gender	Database size
VD	Viennese dialect	HPO	Lower	45-60	M	2:55
VU	Colloquial Viennese	HGA	Higher	60-70	F	3:10
VJ	Viennese youth language	JOE	Lower	15-25	F	2:11

Viennese varieties

Linguistic level	Austrian German Standard	Viennese	Coding level
sound	ə	ɛ	sound
symbol set – phon(em)es	āɐ̯ a	aː / æː / ɛː a / ɔ	lexicon setup
phonology	āɐ̯ 1 # [vāɐ̯l] <i>weil</i> 'because'	ɛː [vɛː]	rules
morphological	<i>pass-te</i> 'would fit' <i>Gläs-chen</i> 'glass dim.'	<i>pass-ert</i> <i>Glas-erl</i>	lexicon transfer
morpho-syntactic	<i>lesen können</i> 'can read' <i>ertrinken</i> 'drown'	<i>derlesen</i> <i>dersaufen</i>	
lexicon: – open class	<i>trinken</i> 'drink' <i>fett, dick</i> 'fat' <i>Kopf</i> 'head'	<i>saufen</i> <i>blad</i> <i>blutzer</i>	lexicon specific
– functional: – articles – pronouns	<i>der</i> 'the' <i>hinaus</i> 'to-out' <i>heraus</i> 'from-out'	<i>d' / da / der</i> <i>ausse</i> <i>aussa</i>	
phrasal: – clitica – infl. compl. – idioms – no preterit	<i>weil du weggehen sollst!</i> 'because you should leave' <i>er ging.</i> 'he went.'	<i>weilst du über d' häuser haun sollst!</i> <i>ea is gangen.</i>	text trans- lation

Figure 1: Levels of representation concerning differences between AT standard and Viennese dialect

Synthesis samples

Computer	Variety		Speaker
	Austrian German	„Hochdeutsch“	SPO
	Viennese dialect	„Wienerisch“	HPO
	Colloquial Viennese	„Umgangssprache“	HGA
	Viennese youth language	„Wiener Jugendsprache“	JOE

Es gibt ja keinen Einheitsdialekt und es kann ihn gar nicht geben, weil jede Wienerin a bisserl anders spricht.

Es gibt Unterschiede nach der sozialen Schicht und nach der Absicht, wie sehr wir Dialekt sprechen wollen.

Wir Wienerinnen müssen nämlich nicht, aber wir können.

Voice development: Speaker selection

- Viennese dialect (VD)
 - actor who came closest to an authentic Viennese dialect speaker although he did produce some stereotypes, which can be seen as beneficial from a listeners point of view
- Colloquial Viennese (VU)
 - actress who had a very natural colloquial speaking style
- Viennese youth language (VJ)
 - pre-selected a specific group defined by age, school-type, gender, and variety spoken within the family

Code	Variety	Speaker	Education	Age group	Gender	Database size
VD	Viennese dialect	HPO	Lower	45-60	M	2:55
VU	Colloquial Viennese	HGA	Higher	60-70	F	3:10
VJ	Viennese youth language	JOE	Lower	15-25	F	2:11

Voice development: Recording

- Conversational speech should be recorded for data-driven speech synthesis of dialect/sociolect
 - dialect is produced as spontaneous, conversational speech
 - no script available
 - hard to annotate automatically
- If read speech is recorded
 - recording script (phonetic transcription) is available
 - automatic annotation (HMM-based forced alignment) is feasible
 - no problem of overfitting
- How to get dialectal speech from read speech
 - use dialectal texts
 - use standard texts with dialect pronunciation, switching between varieties occurs

Voice development: Text selection

- Austrian German recording script is balanced for diphone coverage and prosodic contexts
 - certain word-forms (e.g., preterit) do not exist in dialects
 - certain lexical items do not exist, but have a distinct correspondent
- Filtering of sentences that would be ungrammatical in Viennese varieties. The transcriptions were generated with rule-based methods.
- Ask speakers to read standard text in Viennese dialect
- Thereby we assumed that a good diphone coverage in Standard Austrian correlates with a good coverage in Viennese dialect
- In addition, text scripts from “Viennese” sources in various orthographic encodings were used
 - sentences from comix, poetry, song texts and sentences containing specific Viennese words

Voice development: Phone sets

- Develop base lexica for the phonetic encoding of each variety, which covers the most important and typical words of the respective Viennese variety

Category	Austrian German	Viennese dialect
vowel	a a: (ɔ) ɐ (ɐ:) e: i i: ɔ o:	a a: ɔ ɔ: e e: ɛ ɛ: i i: I o
	u u: y y: ø ø:	o: u u: ʊ y y: ø: œ œ:
di-/monophthong/nasal	æ̃ aõ oẽ	æ: ɒ: ɕ: ɔ̃ ɔ̃ ʊ̃
	(æ̃:) (œ̃:) (õ̃:)	ã: ɔ̃ ɔ̃: î ã õ
r-vocalized	ɐ̃ ɐ̃: ɪ̃ ɪ̃: ʊ̃ ʊ̃: ɔ̃ ɔ̃:	ɔ̃ ɔ̃: ɐ̃ ɐ̃: ɪ̃ ɪ̃: ʊ̃ ʊ̃:
	ũ ũ: ỹ ỹ: ø̃ ø̃:	ʊ̃ ʊ̃: ʊ̃ ʊ̃: (ỹ: ɐ̃) ɐ̃:
schwa	ə ɐ	ə ɐ
plosive/spirant	b d g p t k	b d g β ð ɣ p t k
fricative	f v s ʃ ʒ ʒ x h	f v s s: ʃ ʒ ʒ x h
liqu./nas./glide	ʀ l m n ŋ j	ʀ l l̃ m m̃ n ñ ŋ ŋ j
pause/glottis	'sil' 'pau' ʔ	'sil' 'pau' ʔ

Table 2: Phone sets for Austrian German and Viennese.

Table 2. Definition of phone sets by rules

Rule	P1	P4	P6	P7	P9
merge_eschwa	✓	×	×	×	×
merge_a_aschwa	✓	✓	×	×	×
merge_a_aschwa_l	×	×	×	×	✓
split_Vaschwa	×	×	×	✓	×
split_diphthong	×	×	×	✓	×
rem_V_nasal	×	×	✓	✓	×
neut_mid_v	×	×	×	×	✓
findev	✓	×	×	×	×
rem_findev	×	✓	✓	✓	✓
merge_spirants	×	×	✓	✓	×
despirantize	×	×	×	×	✓
rem_syllabic	×	×	✓	✓	✓
rem_nons_gem	✓	✓	×	×	✓
rem_length	×	×	✓	✓	×
Number of phones	75	76	47	39	66

Voice development: Phone sets

- Encoding all the differences between Viennese dialect and Austrian Standard results in a set of phones that is far too large
 - acoustic models for alignment are based on very sparse data for certain phones
 - diphone coverage is dramatically decreased
- Create reduced phone sets with merge / split and delete rules
- Tests to evaluate phone sets
 - phone-error-rate of letter-to-sound (LTS) rules for different phone sets
 - diphone coverage on a sample of test utterances
 - listening tests
- P9 as winner of the listening test was chosen



Evaluation of phone-error-rate of LTS-rules for different phone sets

Voice development: Spoken dialog system

- Dialog system with 4 personas / synthetic voices that represent a 3-dimensional sociolect space (age, gender, education)
 - (1) Austrian German standard (+/-, male, +)
 - (2) Viennese dialect (+/-, male, -)
 - (3) Viennese youth language (-, female, +/-)
 - (4) Viennese standard German (40+/-, female, +)
- Restaurant scenario derived from evaluation
- Mapping of positive / negative properties to standard / dialect for design guidelines
 - Standard speaker (1) as moderator and help
 - each other speaker has a different type of restaurant associated

Speaker sociolect	Restaurant type
(2) Viennese dialect (VD)	Viennese cooking
(3) Viennese youth language (VJ)	Low prices / cool places
(4) Viennese colloquial (VU)	Luxury restaurants

Release 1.0

- http://data.cstr.ed.ac.uk/festival/festvox_cstr_vd_hanno_multisyn-1.0.tar.gz
Viennese dialect voice (264MB); BSD open source license
- http://data.cstr.ed.ac.uk/festival/festvox_cstr_vd_helma_multisyn-1.0.tar.gz
Colloquial Viennese voice (277MB); BSD open source license
- http://data.cstr.ed.ac.uk/festival/festvox_cstr_vd_julia_multisyn-1.0.tar.gz
Viennese youth language voice (183MB); BSD open source license
- http://data.cstr.ed.ac.uk/festival/festvox_cstr_vd_lex_1.0.tar.gz
Lexical resources and scripts for all voices (Available at 26.5.2010); Academic license
- All links on project website (<http://dialect-tts.ftw.at>) and LREC map by 26.5.2010
- Austrian German voice on <http://www.wien.at>