A Crowdsourcing Smartphone Application for Swiss German: Putting language documentation in the hands of the users

Jean-Philippe Goldman[1], Adrian Leemann[2], Marie-José Kolly[2], Ingrid Hove[2], Ibrahim Almajai[1], Volker Dellwo[2], Steven Moran[3]

[1] University of Geneva, Switzerland, [2] University of Zurich, Switzerland
[3] University of Marburg, Germany
E-mail: Jean-Philippe.Goldman@unige.ch

Abstract

This contribution describes an on-going projects a smartphone application called Voice Äpp, which is a follow-up of a previous application called Dialäkt Äpp. The main purpose of both apps is to identify the user's Swiss German dialect on the basis of the dialectal variations of 15 words. The result is returned as one or more geographical points on a map. In Dialäkt Äpp, launched in 2013, the user provides his or her own pronunciation through buttons, while the Voice Äpp, currently in development, asks users to pronounce the word and uses speech recognition techniques to identify the variants and localize the user. This second app is more challenging from a technical point of view but nevertheless recovers the nature of dialect variation of spoken language. Besides, the Voice Äpp takes its users on a journey in which they explore the individuality of their own voices, answering questions such as: How high is my voice? How fast do I speak? Do I speak faster than users in the neighbouring city?

Keywords: dialect recognition, smartphone application, voice profiling

1. Introduction

Dialect variation is encoded and perceived in everyday language situations. For example, at social events it is quite common to hear exchanges like, "Judging by your dialect, you must be from there, correct?" Although listeners are typically unaware of the underlying linguistic mechanisms involved, they are actively engaging in perceptual dialectology tasks (cf. Preston 1989, Clopper & Pisoni 2004) and they seem keenly aware of dialectal variation. It is interesting then that speakers of different languages seem to identify their dialects with different degrees of accuracy. For Swiss German, Leemann & Siebenhaar (2008) and Guntern (2011) show that naïve Swiss German listeners can accurately recognize a speaker's dialect with a recognition rate of 86%, 74% respectively. However, in other languages, recognition tends to be more difficult: Clopper & Pisoni (2005) report identification rates of only 30-50% for American and British English dialects; Kehrein, Lameli & Purschke (2011) report similar recognition rates for German dialects. One question that underlies the discrepancy of recognition problem is whether the data used in these studies is predictive of accuracy. Recent studies show that dialect recognition is possible via mobile applications (Kolly & Leemann *accepted*) as well as automatic dialect recognition (Biadsy 2011).

This contribution describes the two on-going projects at University of Zurich, leading to two applications called Dialäkt Äpp and Voice Äpp. The main purpose of both apps is to identify the user's dialect on the basis of the dialectal variations of 15 words. The result is returned as one or more geographical points on a map. In Dialäkt Äpp, launched in 2013 (Leeman & Kolly, 2013), the user provides his or her own pronunciation through buttons, while the Voice Äpp asks users to pronounce the word and uses speech recognition techniques to identify the variants. This second app is more challenging from a technical point of view but nevertheless recovers the nature of dialect variation of spoken language.

2. SDS

The Linguistic Atlas of German-speaking Switzerland: *Sprachatlas der Deutschen Schweiz* (SDS, 1962–2003) was collected between 1939 and 1958 with individual interviews in more than 550 places, which represents roughly one third of the existing localities. It was published as 8 volumes including 1500 maps. Various levels of linguistic variations are represented: phonetics, phonology, morphology, and lexical variations.

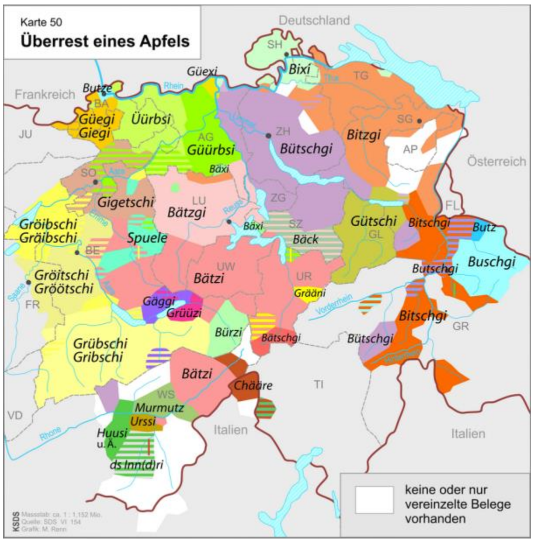
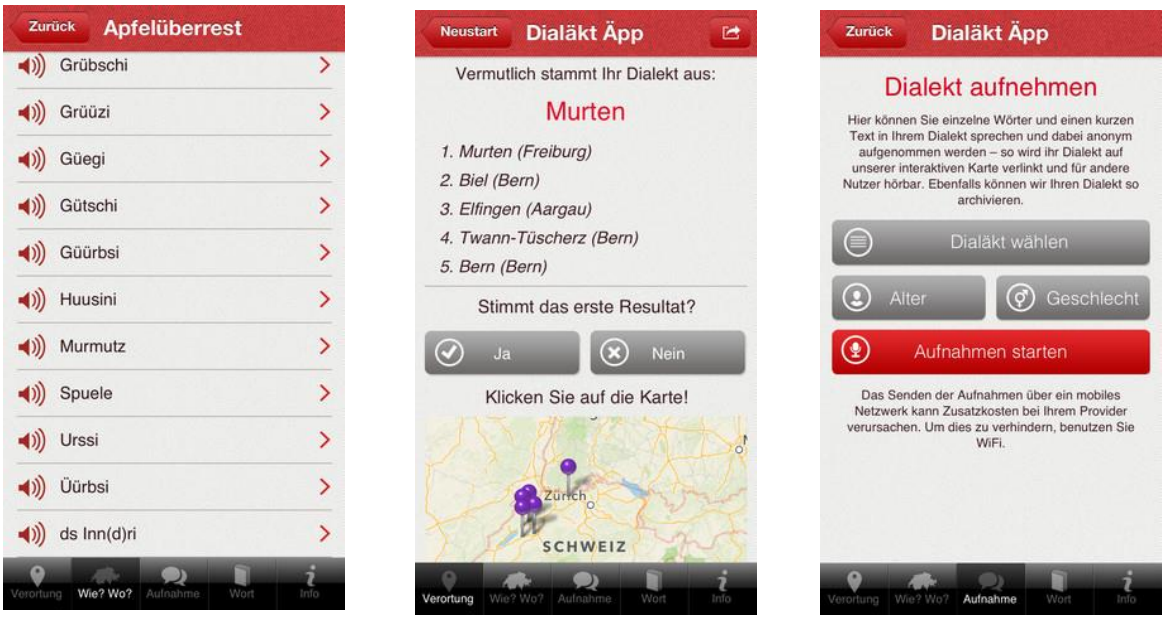


Figure 1. Variant map of 'Überrest eines Apfels', *apple core*.



Vermutlich stammt Ihr Dialekt aus:

Murten

1. Murten (Freiburg)
2. Biel (Bern)
3. Elfingen (Aargau)
4. Twann-Tüscherz (Bern)
5. Bern (Bern)

Stimmt das erste Resultat?

Klicken Sie auf die Karte!

Dialekt aufnehmen

Hier können Sie einzelne Wörter und einen kurzen Text in Ihrem Dialekt sprechen und dabei anonym aufgenommen werden – so wird ihr Dialekt auf unserer interaktiven Karte verlinkt und für andere Nutzer hörbar. Ebenfalls können wir Ihren Dialekt so archivieren.

Das Senden der Aufnahmen über ein mobiles Netzwerk kann Zusatzkosten bei Ihrem Provider verursachen. Um dies zu verhindern, benutzen Sie WiFi.

Figure 2. Screenshots of Dialäkt Äpp : from left to right: dialectal variant choice with buttons, result provided as a choice of five best fitting localities and their corresponding points on a map, and the screen to start recording the user's voice.

For example, the variation of the word 'Bett' (*bed*) make a constrast on the vowel aperture ([bet] for 44% of the places vs [bɛt] 56%). The word 'Tanne' (*fir tree*) has 2 variants distinguishing the length of consonant [n]. The word '*Apfelüberrest*' – apple core – has 39 lexical variants, depicted in Figure 1., such as: *Üürbsi, Güürbsi, Güegi, Spuele, Gige(r)tschi, , Butz, Butze, Buschgi, Butschgi, Bützgi, Bütschgi, Bitzgi, Bitschgi, Bixgi, Bixi, Bätzi, Bätzgi, Bätschgi, Bäxi, Bäck, Gröibschi, Gröibtschi, Göitschi, Gräibschi, Gräutschi, Grübschi, Gütschi, Grüüzi, Gäggi, Chäbi, Grääni, Bürzi, ds, Inn(d)ri, Urssi, Murmutz, Huusini, Ghüüs, Chääruhuusi, Chääre*.

3. Dialäkt Äpp

Sixteen variants (over 15 words) have been chosen among all the maps from the SDS by linguistics experts in order to capture dialectal differences between localities. This set of maximally predictive variations is presented in Table 1. with a number of variants (n1) ranging from 2 to 39. One can notice that the first two variables are about the same word ('Abend' *evening*), dealing with the first vowel and the end of the word. The next columns (n2) represents a slightly reduced number of variants used in Voice Äpp (as explained below), the average number of localities per variant and an estimation of the standard deviation. As can be seen, some words like 'Bett' are well-balanced (with a low standard deviation compared to the mean) as the two variants are dispatched as 44% and 56%, whereas 'Kind' is clearly unbalanced (as one variant out of three represents 97.5% of the localities).

Word	meaning	n1	n2	mean	std
Abend (1st vowel)	*evening*	8	8	62	83
Abend (ending)	*evening*	13	13	40	77
Apfelüberrest	*apple core*	39	21	27	22
Augen	*eyes*	11	7	71	72
Bett	*bed*	2	2	196	53
Donnerstag	*Thursday*	8	7	70	74
fragen	*to ask*	11	7	73	48
heben	*to lift*	3	3	139	92
hinauf	*up*	31	22	26	35
Kelle	*trowel*	5	5	96	55
Kind	*child*	4	3	139	308
schneien	*to snow*	7	5	93	116
spät	*late*	12	9	59	68
Tanne	*fir tree*	2	2	186	149
tief	*deep*	7	7	71	72
trinken	*to drink*	10	5	96	151

Table 1. Sixteen chosen variables and the number of variants from SDS (n1), in Voice Äpp (n2), the average number of localities per variant and an estimation of the standard deviation

The Dialäkt Äpp was launched on March 22, 2013, and has been downloaded around 57'000 times (as of October 14, 2013), i.e., about 2'000 downloads per week on average, with some peaks after media articles. The data recorded by this application contains (a) the (written) choice of pronunciations of 16 variables (over 15 words) by each user who localized his or her dialect and (b) the audio data of the same 15 words by each user who chose to record his or her voice. For (a), our corpus contains data from 39'168 subjects (58% males, 42% females). Most users are found in the cantons (and capital cities) of Zurich, Bern, Basel, Luzern, Aargau, and St. Gallen. 64% of users' pronunciation variants still correspond to the local variant recorded by the SDS (b), the corpus counts 36'617 recorded variants coming from a total number of 2'376 iOS devices (which should correspond roughly to the number of speakers). So, only 4% of the downloads led to a complete recording of the 15 words. 52% of the recordings are from male speakers, 48% from female subjects.

The data recorded by the Dialäkt Äpp has great potential for dialectological as well as forensic phonetic research: It can be used to create new dialect maps and compare them to the maps published in the SDS, to track sound change processes and to create population statistics for a variety of phonetic parameters. It can also be used to compare dialects at the acoustic phonetic level.

4. Voice Äpp

4.1 Adding speech recognition to Dialäkt Äpp

The novelty of this second application is to have the user pronounce the same 15 words with his or her own dialectal variants and to use automatic speech recognition (ASR) techniques to identify the exact pronunciation, instead of buttons on a touch screen as in Dialäkt Äpp. Some difficulties can be overseen as the ASR approach is not error-free, especially through a smartphone application, as the recording conditions may vary a lot (distance from microphone, noisy environment). Also, one must note that this particular task of identifying a dialect variant is not the initial sense of ASR systems as tiny variations have to be distinguished, whereas the speech recognition domain aims at neutralizing such variations and at being rather dialect-independent and speaker-independent. In addition to that, the number of variants of each word is of great importance. For example, the word Bett (*bed*) has only 2 variants (/bet/ and /bɛt/) whereas "Augen" (*eyes*) has as many as 7 dialectal variants. Thus the latter word is very discriminant in Dialäkt Äpp (as long as it is correctly entered with buttons), but the task is much more difficult for an automatic speech system, and an erroneous recognition will lead to a wrong dialect localization. As the voice recognition approach is not as reliable as the selection with buttons, the algorithm has to be slightly modified to take this uncertainty in account.

For this, we transcribed the recordings from the DialäktÄpp. So far five words of the 2'376 users who recorded their voice have been transcribed. A comparison between these data and the SDS data led for simplification to a slightly reduced number of variants for each word. Differences can be seen in Table 1 between columns *n1* (SDS) and *n2* (Voice Äpp). Then, 5 full-word speech recognizers have been individually trained on each of the words.

A test set of 100 speakers, external from the training set, has been selected for variant evaluation. The Table 2 shows the percentage of recognition rate for the 5 words ranging from 80% to 99% of accuracy individually.

Word	n2	% accuracy	#items
'Apfelüberrest' *apple core*	32	80.37	1180
'Bett' *bed*	2	86.92	331
'Fragen' *to ask*	7	83.18	1719
'Kind' *child*	3	99.07	410
'Tanne' *fir tree*	2	89.72	148

Table 2. Number of variants for the 5 words, percentage of recognition accuracy, and number of items in the training sets

The next step will be to gather these 5 word-recognizers into a dialect decision algorithm in order to test complete dialect identification.

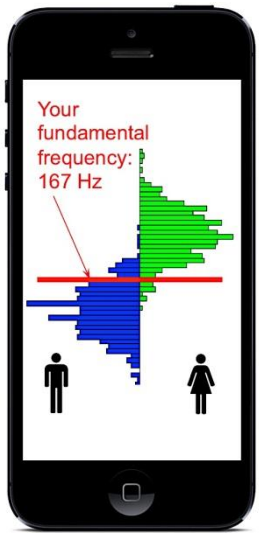
4.2 Voice profiling

The second main function alongside the localization of the user's dialect is the voice profile. In this part, the user gets to know characteristics of his or her own voice and he or she learns about speech production and perception in a playful way.

When starting the application, the user is asked to record a sentence in his or her dialect and indicate age and sex. This sentence is used as a basis for the voice profile. Next, a number of menus and submenus allow the user to explore different aspects of speech, namely pitch, speech rate, articulation, auditory perception and visual perception.

4.2.1 Pitch

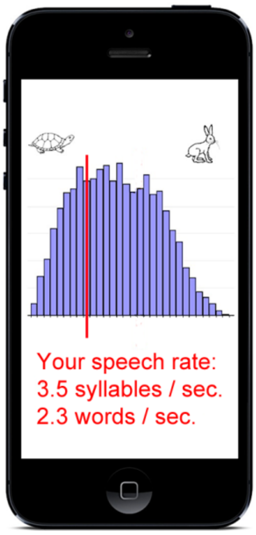
The fundamental frequency of the user's sentence is calculated and displayed in a histogram representing the distribution of the fundamental frequency of all of the previous users of the application. The user can thus compare his pitch to the other users.



Other functions within the category of pitch include the following: The user can listen to what his or her own sentence could sound like if he or she was a person of the opposite sex; the user can try to reproduce the sentence in a very high or very low pitch; and a video of vibrating vocal folds can be watched.

4.2.2 Speech rate

The speech rate of the user's sentence is calculated and displayed in comparison to the other users' speech rate.



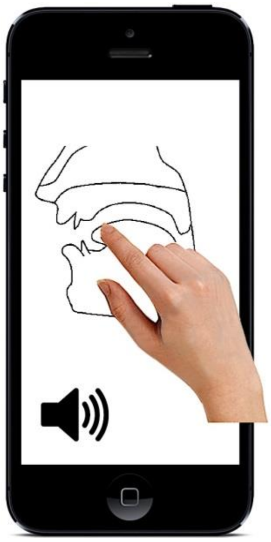
Your speech rate:
3.5 syllables / sec.
2.3 words / sec.

4.2.3 Articulation

The user can learn about sounds and their phonetic notation. Upon clicking on an IPA symbol a sagittal cut is shown and the corresponding sound is played. Moreover, in an interactive sagittal cut the user can move the position of the tongue and listen to the sound which corresponds to the position of the articulators he or she just created.

4.2.4 Auditory perception

The user can listen to what his or her sentence sounds like to a person with a hearing impairment or to a person with a cochlear implant.



4.2.5 Visual perception

The user is shown a video of a person with the lip movements of [ga ga] but the simultaneously played sound is [ba ba]. After the user ticks off whether he heard [ba ba], [da da] or [ga ga], an explanation is provided as to why most people perceive [da da] (known as McGurk effect (McGurk & MacDonald 1976).

Furthermore, a video of a person speaking in a very noisy environment is played. At first the user does not see the speaker's mouth, later he does. This serves to demonstrate that under adverse listening conditions it is easier to understand a person when visual cues provide additional information called cocktail party effect (Handel 1989). A small quiz will allow the user to test his or her own lip reading abilities.

The Voice Äpp application should be as interactive as possible, allowing the user to learn about characteristics of his or her voice and about speech production and perception in general by "playing around". Pictures and graphs are preferred over text. At the same time, if the user wants to, he or she has the possibility to obtain more details on a certain topic by activating a screen page containing background information.

5. Conclusion

The upcoming steps are twofold. 1. The dialect localization will be achieved with the transcription of further existing speech material (either from the Dialäkt Äpp or additional recordings), a global training and a decision algorithm, currently under development. 2. The voice profiling needs to be tested with its first users and their feedback will be crucial for future enhancements.

6. Acknowledgements

The project Swiss VoiceApp – Your voice. Your identity is funded by the Swiss National Science Foundation (SNSF); funding scheme: Agora; grant number: 145654.

7. References

Biadsy, F. (2011) Automatic Dialect and Accent Recognition and its Application to Speech Recognition. PhD, University of Columbia.

Clopper, C.G., D. Pisoni (2005). Perception of dialect variation. In: Pisoni, D., R.E. Remez (Eds.), *The Handbook of Speech Perception,* Oxford: Blackwell, 313-337.

Dellwo, V.,Leemann, A., Kolly, M.-J., (2012) Speaker idiosyncratic rhythmic features in the speech signal, IN Proceedings of Interspeech-2012, PP 1584-1587.

Ferragne, E., & Pellegrino, F. (2007). Automatic dialect identification: A study of British English. In Speaker classification II (pp. 243-257). Springer Berlin Heidelberg.

Guntern, M. (2011). Erkennen von Dialekten anhand von gesprochenem Schweizerhochdeutsch, *Zeitschrift für Dialektologie und Linguistik* 78.2, 155-187.

Handel, S. (1989). Listening. An Introduction to the perception of auditory events. Cambridge, Mass.: MIT Press.

Kehrein, R., A. Lameli, C. Purschke (2010). Stimuluseffekte und Sprachraumkonzepte. In: Anders, C., M. Hundt, A. Lasch (Eds.), "*Perceptual dialectology". Neue Wege der Dialektologie*. Berlin/New York: de Gruyter, 351–384.

Kolly, M.-J., Leemann, A. (accepted). Dialäkt Äpp: communicating dialectology to the public – crowdsourcing dialects from the public. In: A. Leemann, M.-J. Kolly, S. Schmid, V. Dellwo (Eds.), Trends in Phonetics in German-speaking Europe, Bern/Frankfurt: Peter Lang.

Leemann, A., M.J. Kolly (2013) Dialäkt Äpp. https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8

Leemann, A., Siebenhaar B. (2008). Perception of Dialectal Prosody, in *Proceedings of Interspeech 2008*, pp 524-527

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 746-748.

Preston, D. (1989). Perceptual dialectology: nonlinguist' s views of areal linguistics. Dordrecht: Foris.

SDS *Sprachatlas der deutschen Schweiz*. (1962-2003). Bern (I-VI), Basel: Francke (VII-VIII).