Talebob - An Open-Source Tool for Training Spontaneous Speech

Peter Juel Henrichsen
DanCAST - Danish Center for Applied Speech Technology
Copenhagen Business School
Denmark

Abstract

In this article we present the Danish speech trainer Talebob ("Speech Bob"), an interactive CALL[1]-tool developed at the Danish Center for Applied Speech Technology. Talebob is intended for use in language classes in a vast and thinly populated part of the Nordic area where IT-based teaching materials are in high demand. The tool allows the pupil (aged 9 to 14) to train his or her spoken Danish concentrating on pronunciation aspects such as speech tempo, prosodic contour, and articulation. Having tested Talebob in four countries, we report on our technical, phonetic, and didactic results. The main purpose of the present article, however, is to introduce the Talebob developer's workbench allowing foreign language teachers in any country to import and localize Talebob for use in his or her own classroom.

1 Introduction

In the European community, the Danish language with its 5-6 million speakers is usually considered *small* (not to say insignificant). To many Europeans it comes as a surprise that Danish is taught in school as an obligatory foreign language in least three independent parts of the Nordic area, Iceland, Greenland, and the Faroese Islands. For historical reasons, Danish became – and still is – a lingua franca across this huge area, and a large proportion of its inhabitants, at some stage in their education, still end up in Denmark. Danish language skills are therefore still in high demand. However, not all aspects of modern Danish are taught equally carefully.

Most Danish classes in, say, Greenland concentrate on reading, supplemented by exercises in writing, grammar, vocabulary, and (to a lesser degree) phonetics. Almost no attention is given to the language style that any immigrant faces immediately upon his or her first arrival in Denmark: the vernacular.
Danish spontaneous speech is as challenge to any new-comer (even Swedes and Norwegians with languages much more similar to Danish), with its extraordinarily high syllable-per-second ratio and is pervasive phonetic lenitions, reductions, and assimilations. Consider a few examples.

"det er jo ikke noget at snakke om"
(8 lexical syllables, full vowels underlined)
[djoJgnåD:snagCm][2]

(4 phonetic syllables, full vowels underlined)

This often heard phrase (literaly: *that is nothing to talk about*, meaning: *it's not a problem*) is routinely uttered in four phonetic syllables only, and with a highly predictable prosodic contour. If pronounced in accordance with the productive rules of Danish phonetics, reproducing *all* of the phonological vowels (as a typical rule-based TTS voice does), this phrase would probably be perceived by the native Dane as a composition of several independent semanitic units in various relations, a speech act (*snakke om*), a predicative modifier (*jo*), and a negated quantifier (*ikke noget*), in short, a fully fledged proposition to be compositionally evaluated.

Consider an other example, "tak skal du have" (literally: *thanks shall you have*, meaning: *thank you*), along with its highly idiomatic pronunciation patterns.

[1]CALL = Computer Assisted Language Learning
[2]Here (and in the following) phonetic renderings are shown in SAMPA compliant format, cf. http://www.phon.ucl.ac.uk/home/sampa/danish.htm

[t'Agsgaduh,a:?] unmarked-polite, mildly grateful
[t'Agsgaduha] impressed (no gratitude involved)
[tAgsgad'uh,a:?] repulsed, sullen (anti-grateful)

Since many Nordic language teachers had their didactic training in times where skills in scholarly Danish (slow and well-articulated) were in higher demand, and since most have had little or no contact with native Danes since then, new didactic methods and new teaching materials reflecting the current pronunciation habits will be needed for preparing the pupil for the language encounter. Indeed, the majority of the Nordic pupils we have interviewed about their experience with meeting the Danes, report that they did not understand them at all, and did not feel understood either when trying their own luck. Many felt disappointed, believing that they had a good conduct of Danish.

For these reasons we[3] decided to develop a modern, internet-based tool for training (aspects of) 'free-style' spoken Danish, suited for school children in 3rd to 7th grade.

Talebob is designed for practicing the pronunciation of frequent phrases, often rich in function words (pronouns, connectives, adverbs and prepositions). As mentioned, their pronunciation patterns are typically highly conventionalized and are often in conflict with the general and productive rules of Danish pronunciation.

In the following, we first present Talebob in its current version and then reflect on how to develop the tool further. Sections 2-5 cover the technological and linguistic aspects of Talebob's design (front-end, back-end, and system architecture). In section 6 we reflect on various linguistic aspects of Talebob, in current and future versions. Section 7 is devoted to the newly published portation toolkit. We conclude in section 8 with some remarks on Talebob (and interactive language learning tools in general) as an approach to screening large populations of pupils.

Example phrases are quoted in Danish and (being highly idiomatic) translated only when necessary.

2 Talebob as a CALL tool

Talebob is a tool for computer-assisted language learning (CALL), and it can be seen as a technically updated continuation of the classic language lab. Many readers will probably remember from their school days the setup with study booths equipped with a cassette deck for recording and playback, enabling oral communication with the language teacher on a one-to-one basis. The language lab (e.g. Thorborg 2003, 2006) stimulated the pupil's spoken language production and in this respect was a huge improvement over L2 exercises based on rehearsed dialogues. Of course the attention from the teacher was a scarce resource, and each pupil could not expect more than a few minutes of personal instruction during a lesson.

One of our main goals with Talebob is to take the language lab a step further towards interactivity such that each language production will yield an informed comment, either an appreciation or a constructive correction. In other words, Talebob should give the pupil a feeling of being heard.

3 Talebob's front-end

School children are used to computer games with a visual side approaching virtual reality. Rather than competing on graphics we wanted to attract our users through a carefully designed interactivity offering meaningful replies on all contacts. Talebob should thus behave as an attentive listener and competent evaluator.

The Talebob challenge consists of 30 tasks, each focused on a specific Danish phrase such as greeting formulae (*godmorgen*), common requests (*gi'r du en kop kaffe?*), and emotional expressions (*er du rigtig klog?!*). Common to such phrases is that their communicative effects may change radically with the smallest twists of the pronunciation. An inconspicuously looking phrase like "tak skal du have" (*thank you*) may be perceived as being ironic, impressed, tired, cordial, hateful, or just plainly informative depending on subtle prosodic modifications (e.g. changing the relative weight of the main stresses slightly). Being able to control such details is an intrinsic part of one's L1 competence, but is often difficult for L2 learners to acquire. Talebob allows the pupil to repeat each phrase as many times as needed, informed by Talebob's feedback. The phrase prompts are produced by a native speaker aiming for an 'ecological' pronunciation that no Dane would object to.

[3]The author, in close collaboration with Iceland University (above all prof. Auður Hauksdóttir, see Acknowledgements). Talebob is a component in the Taleboblen parcel of teaching materials (cf. taleboblen.hi.is) to which researchers from Stockholm University and Gothenburg University have also contributed significantly.

For each Talebob-task the pupil

1. selects a phrase,
2. listens to the phrase prompt (using the Lyt-Til-Frasen button),
3. reproduces the prompt orally (using Optag/Stop buttons for recording), mimicking it closely wrt. articulation, prosody, and tempo,
4. compares prompt and own production auditorily (pressing Lyt-Til-Optagelsen),
5. repeats steps 2-4 until entirely satisfied, then presses Send for evaluation,
6. consults the returned Talebob comment (either a success message sending the pupil to the next task, or a try-again advising the pupil how to improve)

Pressing Send invokes the Talebob acoustic analyzer, returning a smiley, either happy, neutral, or sad. With a happy smiley **:-)** the pupil has completed the task and may continue with the next phrase. Level-1 is done when the first five tasks are completed, level-2 has ten tasks, and level-3 fifteen. The phrases are ordered progressively, from single words and simple phrases in level-1 (*godmorgen*, *værsgo!*), frequent idioms in level-2 (*hvordan går det?*, *tak i lige måde*), to more expressive phrases in level-3 (*det siger du ikke?*, *hellere end gerne!*). When all tasks in level-3 are done, the Talebob challenge is passed.

Talebob's front-end is illustrated in fig. 1-3.

4 Talebob's back-end (acoustic analysis)

The two sound files submitted (with the Send button) are evaluated in the Talebob back-end application. The acoustic analysis compares the prompt version (P) and the user's own production (U) sampling both files for F0 (pitch in Hz) and INT (intensity in dB), being unanimously considered as the most relevant parameters for acoustic-phonetic evaluation, both relating directly to phonetically features like stress, tone, sonority, occlusion, etc.[4] The linguistic evaluation is focused on the concordance of P and U wrt. speech tempo, global prosody, and articulation.

The speech tempo factor (*STF*) is determined as the ratio of durations for P and U,

$$STF = \text{duration}(P) : \text{duration}(U)$$

STF is calculated from INT data. First the zero level for INT in U is estimated, corresponding to 'no speech' in the given signal (this calibration can be tricky, especially for noise-prone samples, and is always a matter of heuristics). Then the zero level (0 dB after calibration) is used to delimit the speech production in U. By definition the optimum value for *STF* is 1.0, and productions approaching this value will trigger the comment "Meget fint taletempo" (*excellent speech tempo*). Lesser or greater values return instructions to speak faster or slower, respectively.

Prosody and articulation analyses are based on F0 measurements. Only the sonorant parts of P and U are sampled - that is, the segments of the speech signals where a pitch value can be meaningfully estimated, thus excluding obstruent sounds and moments of silence (e.g. between words). All frequency data are stored as logarithmic values (more convenient for statistical use). Many of Talebob's users are children, and their speech productions will often be higher-pitched than the phrase prompt on average. This global difference in pitch is of course irrelevant to the Talebob evaluation, so the F0 dataset for U is normalized (each sample multiplied with a derived constant) equalizing the average pitch of U and P.

[4] F0 and INT are measured using the Praat toolkit (www.fon.hum.uva.nl/praat), window size 5 ms, filter settings = *Pitch (ac)... 0.005 75 15 yes 0.03 0.45 0.01 0.4 0.14 600; Intensity... 75 0.005 yes.* We also experimented with HNR (harmonicity-to-noise ratio) and various spectral filterings, but found them to be too noise sensitive. Classrooms are not quiet places!

Figure 1. Screenshot (excerpt) from Talebob task-page, level 2, with one phrase passed.

Figure 2. Screenshot (excerpt) from Talebob return-page, level 2, not-passed.

Figure 3. Screenshot (excerpt) from Talebob return-page, level 2, passed.

After these preparatory steps, the prosodic evaluation is done. The calculation is based on 10 qualified datapoints for each (normalized) dataset U and P, in a procedure best explained by an example. Say 130 valid pitch samples were derived from P; the first datapoint for P (call it $f_{1,P}$) is then derived as the mean value for the first 13 samples; the 2nd datapoint ($f_{2,P}$) for samples 14..26, et cetera, up to ($f_{10,P}$) and ($f_{10,U}$). Finally the prosodic deviation (*ProsDev*) of U wrt. P is calculated by summation of 'errors',

$$ProsDev = |f_{1,P}-f_{1,S}| + |f_{2,P}-f_{2,S}| + .. + |f_{10,P}-f_{10,S}|$$

This particular *ProsDev* formula was designed to meet two special requirements. Firstly it abstracts away any temporal incongruities between U and P (already addressed by the *STF* score); secondly it copes well with the unpredictable number of valid F0 samples for U (sometimes as few as 15-20 for short speech productions in noisy surroundings, while P may produce 3-4 times more), preserving commensurability. For low *ProsErr* values, Talebob returns a praising comment "Dit tonefald er fint", and otherwise an instruction how to improve, e.g. "Prøv at tale mere livligt" (*try speaking more lively*).

The articulation is evaluated (*ArtEval*) along the same lines, but focusing on local incongruities rather than the phrase as a whole. First 30 qualified datapoints are derived following the procedure above, using numerical interpolation if necessitated by data sparseness. Error analyses (calculated as for *ProsDev*, mutatis mutandis) are done for datapoints 1..10, 11..20, and 21..30,

$$\text{ArtEval(a,b)} = \sum_{n=a}^{b}\left(F_{n,P} - F_{n,U}\right),$$

F being is the 30-point dataset (otherwise as *f* above). The results for *ArtEval*(1,10), *ArtEval*(11,20), and *ArtEval*(21,30) represents the first, middle, and last part of the utterance as reflected in the returned comments: "Prøv at tale tydeligere i de første/midterste/sidste ord" (*try to speak more clearly in the first/middle/final words*), a somewhat vague instruction perhaps, but faced with the impatience and limited vocabulary of pupils we had to prioritize didactive effect over descriptive accuracy.

Summing up, feedback from Talebob consists in three comments, one for each of the evaluation criteria (tempo, prosody, and pronunciation), and in addition a smiley representing the overall performance. The *happy* smiley ('task completed') is given when each of the three evaluation results has met a (pre-set) acceptable limit, the *sad* smiley is given if none of the limits are met, and the *medium* smiley otherwise.

See the discussion below on the linguistic relevance and scientific testability of the Talebob acoustic-phonetic design.

4.1 An example - phrase "hej med dig"

The graphs in fig. 4 and 5 both cover the phrase *hej med dig* in three speech productions, (*i*) the prompt, (*ii*) an Islandic pupil (boy, 7th grade) on 2nd attempt, and (*iii*) same pupil on 5th attempt. Notice that INT graphs are continuous, intensity being defined everywhere, while F0 graphs are interruped at unvoiced passages (e.g. the stopped [d] in *dig*).

The huge difference in speech tempo between 2nd and 5th attempt is easily appreciated in fig. 4. The very slow tempo in #2 (2nd attempt) triggered the Talebob comment "Du taler alt for langsomt" (*you speak much too slowly*); the pupil sped up and - as seen - eventually matched the prompt's tempo in #5. His pronunciation had also become more fluent, without the unwarranted separation of *hej* and *med* (cf. the INT dip around *t*=0.45" in the #2 graph, absent from both #5 and the prompt). Concerning the prosodic contour, notice that the F0 envelope for #2 and #5 (cf. fig. 5) both match the prompt quite closely when abstracting away from the different tempi: two stable pitch inclinations with an intervening resetting, corresponding to the two stress groups in the (most common) Danish pronunciation. Consequently, *ProsDev* is relatively low in both cases, having Talebob praise the pronunciation in both cases: "Meget fint tonefald" (*very good tone-of-voice*). At the same time, though, the *ArtEval*-based analysis shows a 'lack' of pitch modulation in #2 (preceived as mumbling, and producing a relatively poor *ArtEval* value), in this case triggering the comment for #2: "Prøv at tale tydeligere" (*try to pronounce the words more clearly*). Through his next attempts, the pupil improved his pronunciation gradually, and by #5, the *ArtEval* value passed the accept limit, allowing Talebob to issue a happy smiley (notice though in fig. 5 that the pitch range is still somewhat limited for #5).

5 System architecture

The Talebob development had three phases. First an appropriate set of phrases was selected and recorded, largely recycling materials and selection criteria from earlier CALL projects including Allwood et al (2005), Selsøe et al (2004), Henrichsen (2004, 2004b, 2014). Then the back-end was programmed and tested (main programs written in Perl using the standard open-source modules only, enhanced with Unix system calls). The front-end, however, presented us with an unexpected challenge. Nobody could update us on the IT situation in West-Nordic schools, neither for hardware, software, operating system, local IT-assistance, or even internet connectivity. Yet we did not want any potential user to go down on equipment. Also we did not want to preclude any working places. Some pupils prefer to train in the privacy of their home while others like to share. We did not want to force any limitations on the user on purely technical grounds. This led us to consider three front-end/back-end architectures (presented as *A1*, *A2*, and *A3*).

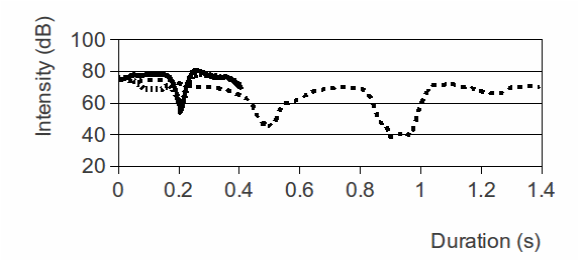

Figure 4. Phrase "hej med dig", intensity data; prompt (solid line), Icelandic pupil's 2nd/5th attempt (dispersed/close dots)

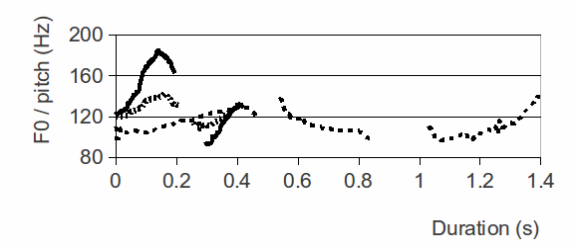

Figure 5. Phrase "hej med dig", pitch data; prompt (solid line), Icelandic pupil's 2nd/5th attempt (dispersed/close dots)

A1. Stand-alone (program installed on user's own hardware: pc, tablet, or smartphone)

PRO:

- Independent of internet connectivity
- Quick query-response cycle

CON:

- Programming/maintenance of back-end for a range of unknown hardware is demanding
- Technical support (from developer to pupil, teacher and/or local IT helpdesk) is hard due to physical and time-zone distance
- Monitoring of users' performance and progress is difficult
- System updates are hard to communicate

A2. Browser-based

PRO:

- Contacts between users and server can be logged (easier maintenance & development)
- Developers can make performance data available to teachers and others online
- Browser-based front-end using HTML5 and CSS is (fairly) hardware independent

CON:

- Stands or falls with user's connectivity
- 100% server uptime is mandatory
- HTML5 audio, especially for recording, is currently not fully supported in all browsers

A3. Internet-based, but dedicated front-end

The advantages are the same as for *A2*, and in addition the HTML5 problem can be avoided. Also we do not need to instruct users to download this or that internet-browser. The main hurdle being that users have to install a dedicated program prior to their first positive Talebob experience.
Even if *A2* seemed to us to be the best alternative overall, we settled on *A3* for practical reasons. Many potential users are Explorer fans and did not care to install a new browser with better HTML5 support, such as Chrome, Firefox, or even IE 9+.

As the developer team had some experience with Unity4 (www.unity4.com), in particular its strong audio support and graphics drivers, we settled for this programming workbench. Unity4 is freely available (in the open-source version) and so does not compromize Talebob as a shareable application.

Unity4 programs compile to all common operating systems (even older versions) including Linux, Mac, Win, Android, etc. The flip side of the coin is that potential Talebob users have to download an executable (via Dropbox, as explained in the Taleboblen homepage, www.taleboblen.hi.is), unzip it, and invoke it using their own operating system. Simple as these procedures may be for skilled IT-users, they showed to be problematic for many language teachers and even local IT-helpdesks. We intend to launch a purely browser-based Talebob-version in the near future, as a supplement to the current version.

For an interesting discussion on CALL design principles for tools training spoken language, see Appel et al (2012). González (2012) and Mbah et al (2013) have experimented with minimalistic CALL applications for English teaching.

6 Talebob meets the world

Before launching our test programme in Iceland, Greenland, and the Faroese Islands we wanted to assess Talebob's competence as a Danish language teacher, so we evaluated Talebob with a panel of native Danish speakers (18 pupils aged 9-18), in surroundings chosen to match the typical Talebob user's (school, car, living room). 16 out of 18 panel members completed the 30 phrases in less than 50 attempts, meaning that most tasks were completed on the first attempt. This seemed to be a satisfactory result.

For comparison, our current log of L2 users at the time of writing shows an average of 84 attempts for the Talebob challenge as a whole (2.80 attempts per phrase), with a global best-score of 55 attempts. Danes and non-Danes thus seem to be clearly distinguished, suggesting that Talebob's automatic feedback is linguistically non-arbitrary as well as didactically useful.

6.1 The case of Iceland

Table 1 summarizes all contacts made to the Talebob back-end during our (still ongoing) test period. For technical and practical reasons, Greenland and the Faroes have only been able to access Talebob systematically for a considerably shorter time than Iceland. We therefore have to postpone cross-country comparisons to a later paper.[5] The pupils taking part in the experiment were not urged to finish the Talebob challenge. They were simply invited by their teacher to try it out. It's therefore interesting to notice that approximately half of the users who have taken up the Talebob challenge (i.e. passed at least one phrase task), do finish the course as well. In other words, we don't see signs of 'early fatigue'.

When consulting the performance data, we see that level-1 phrases took 2.64 attemps to pass on average, level-2 took 2.54, and level-3 took 3.48. As level-3 puts the user under much heavier demand (15 several-word phrases, compared to level-1's 5 very short phrases), we conclude that pupils, in general, are not scared off by the harder struggle. Out of 19 pupils entering level-3, almost 70% completed the level as well. This is an encouraging result, convincing us that Talebob - even in it's earliest version, with crude graphics, canned messages, an adult prompt voice, and no personalization at all - can be appreciated as a fun and meaningful challenge by young children used to the far more advanced interaction of computer games.

[5]The cross-country study could be an interesting one given the extremely different attitudes towards Danish as an L2 encountered in the West-Nordic area. Running a risk of premature generalization, we observe that Greenlandic pupils are highly motivated learners (being heavy users of Danish media) as opposed to the Icelandic children who may have an easier time pronouncing the Danish sounds, but are generally much less motivated anyway (Iceland being in some respects more culturally self-sufficient). Faroese children don't seem to question the necessity of learning Danish at all (many of them preparing for studies in mainland Denmark).

Log-data (TB=Talebob)	**All**	**Iceland**
TB contacts	2508	1888
TB phrase evaluations	2203	1773
Level-1 commenced	39	27
Level-1 passed	30	23
Level-2 passed	24	19
Level-3 passed	16	13
Smiley-1 (*happy*)	738	571
Smiley-2 (*medium*)	1355	1123
Smiley-3 (*sad*)	110	79
TB-eval. per Smiley-1	2.99	3.11

Table 1. Log-data for Icelandic users as per 18/12 2013. Column 'All' includes Faeroese and Greenlandic contacts.

6.2 Talebob as a scientific enterprise

Our current evaluation scheme (based on *STF*, *ProsDev*, and *ArtEval*) has worked well, providing a useful compromise between linguistic precision and communicable (age-appropriate) advise. However, we are aware that this particular setup has not proved itself in a strict scientific sense. Maybe different formulae or new scoring procedures would allow even more useful feedback from Talebob. For example, we suspect that *ProsDev* and *ArtEval* definitions based on standard deviation rather than numerical distance may allow more specific corrections. New batteries of formulae are constantly being tested - still without this being driven by ideal linguistic criteria, but rather as a pragmatic and feedback-informed activity.

Actually, it's not clear to us that an 'ideal' configuration could be obtained at all. The most effective evaluation procedures, from a didactic point of view, would not rely solely on ideal linguistic criteria, but include the personal profiles of the pupils (degree of motivation, prior knowledge of Danish, own first language, general IT-experience, and more).

6.3 Productive expressivity

Talebob is, of course, a low-knowledge system with very little in-built language competence. Inspired by the special focus of NLP4CALL we reflect upon how to induce an amount of linguistic 'intelligence' in Talebob without compromising the low-knowledge style tenet (we'll return to this point shortly).
After having passed level 3, users should feel comfortable with the Talebob feedback cycle. The new prosodic awareness could be developed further by having the user engage in a 'real' dialogue, exploring a kind of interactivity where the choice and production of a phrase (as opposed to another realization of the same lexical word sequence) have direct consequences for the continuation of the game (and score!).
By way of illustration, consider again the phrase *tak skal du have* repeated here for convenience.

p1. [t'Agsgaduh,a:?] polite, mildly grateful
p2. [t'Agsgaduha] impressed, shocked
p3. [tAgsgad'uh,a:?] repulsed, sullen

As opposed to the game levels 1-3, Talebob now takes the initative presenting an assertion among a1-a3 (randomly chosen).

a1:- her er din kaffe
(your coffee, here you are)
a2:- jeg har lige set en trafikulykke
(I just witnessed a traffic accident)
a3:- du skal da bare betale den ødelagte dør
(*why don't you just pay for that broken door*)

The user responds to the assertion by selecting one of the prosodic renderings P1-P3 of the target phrase and then uploads his speech production.
Talebob performs an acoustic comparison between the user input and the canned versions, deciding the closest match and, hence, how to continue the conversation in a coherent manner.
Coherent discourse

T:- her er din kaffe (*here is your coffee*)
U:- tak skal du ha' [neutral-polite, mildly grateful]
T_a:- bruger du mælk og sukker? (*milk or suger?*)
or
T_b:- var der andet? (*you want anything else?*)

Anomalous user input

T:- her er din kaffe (*here is your coffee*)
U*:- tak *SKAL* du ha'! [impressed/chocked]
T*$_a$:- er der noget galt? [is *something wrong?*]
or
T*$_b$:- gør du nar af mig? [*are you making fun?*]

Likewise for the other predictable dialogue paths. Probably only a subset of the phrases included in the current Talebob will be suitable for this new "stimulus-response" scheme, calling for new selection criteria in the compilation of the phrasicon (phrase selection). Single- and multi-word interjections ('ja', 'nej', 'nå'[6], 'okay', 'klart', 'hold kæft', 'er det sandt?', etc) immediately spring to mind. As a side-effect of this construction work, we -- the linguists -- will probably also learn our own language better!

We consider using TTS for presenting the priming assertions, adding still more realism to the dialogue training. We will need a synthetic voice giving us full prosodic control. For this reason we opt for a diphone voice, since the (more modern) unit-selection based voices typically achieve their naturalness by sacrificing the prosodic control of the output. With the TTS-enhancement, one could have even the priming assertion itself change its triggering potential (i.e. the adequate response) as a function of it's prosodic contour alone.

7 The portation tool kit

As discussed above, there is nothing intrinsically 'Danish' about Talebob; the acoustic analysis and scoring procedures have no language-specific parts. Indeed, any user utterance with a pitch envelope similar to the reference utterance would achieve a high score, regardless of the lexical content of the utterance. This can be seen as a strength or a weakness in a broader NLP perspective, and indeed our academic discussion partners have expressed a wide range of opinions about this. Suffice it to say that we have not, until now, encoutered any 'cheating' among pupil users, rather the opposite: judging from our own evaluation of the recorded sound productions, all pupil users without exception appear to have worked on improving not only their prosodic performance (which is monitored by Talebob), but also their phonetic accuracy (which is not). This benign placebo effect is, in our view, an important observation in its own right, sharing in effect the evaluation burden between the CALL tool (which can never compete with a professional language teacher anyway) and the learner himself (who may not even be aware of his self-monitoring). In order to quantify the placebo effects, we would need to perform a controlled experiment with two user groups, one using a mock-version of Talebob producing random feedback, and one control group using Talebob as is. We have not performed such an experiment, but it might be an interesting one.

As said, the value of low- or no-knowledge CALL tools is a controversial issue. However, in one respect, Talebob's linguistic ignorance is an undisputable advantage. When porting Talebob to new L2 teaching scenaria, hardly any software modifications will be needed, only an editorial process of selecting 30 (or more) suitable phrases followed by a recording session with one or more native speakers with a flair for 'ecological pronunciation'. The technical integration of these materials is fairly trivial (though some languages may require slight changes in the acoustic setup). In this respect, Talebob's simplistic speech evaluation differs from the technologically far more sophisticated CALL tools for L2 conversational training available in the market, such as Cooori (www.cooori.com), Wang (2011), de Vries (2014), and Mirzaei et al (2014), all including a fully-fledged ASR component (automatic speech recognition).

[6]The many semantic facets of the Danish interjection 'nå' [n'C] is ingeniously protraited in the famous song, by Poul Henningsen

We have recently published a tool kit allowing any L2 teacher, perhaps with some general IT experience, to compile a localized Talebob version for use in his own classroom. As illustrated in figure 6 below, the necessary activities are concentrated on (i) compiling the phrasicon (based on a manual of selection criteria), (ii) producing the speech prompts (in accordance with certain pronunciation principles), (iii) adjusting the Commentary (therepertoire of eventual feed-back messages), (iv) generating the two essential executives (the front-end and back-end), (v) installing the *BE* (back-end application) on one's local web-server and, finally (vi) distributing the *FE* (front-end aplication) to the pupils or other end-users.

The toolkit, together with the necessary manuals, can be downloaded from the project URL.
http://lab.homunculus.dk/Talebob_portation_ toolkit
In case the link is down (the server unfortunately has not been fully maintained since the project budget ran out), please do contact the author for further information.

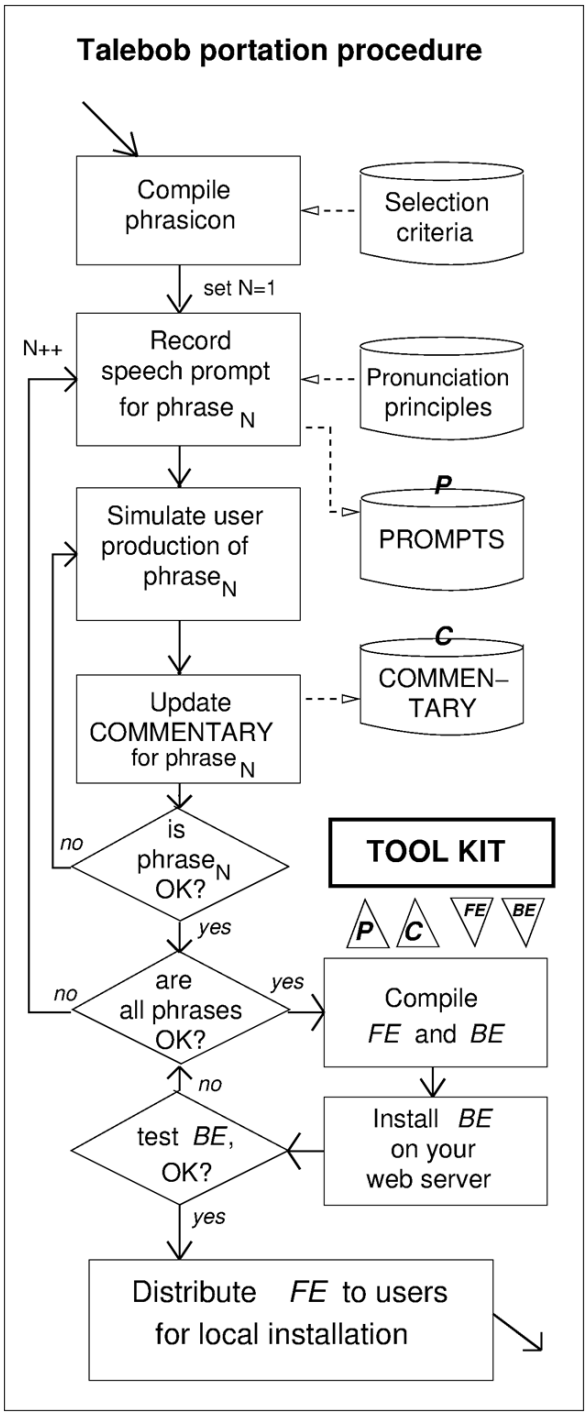

Figure 6. Using the Talebob Portation Tool Kit.

7.1 CALL-based monitoring

In this concluding section we touch on CALL tools for societal use in a broader perspective (with Talebob as an example) as a means of gathering data not only relevant to didactic practices and research, but to basic linguistic research as well, and even (potentially) to political bodies, providing them with quantitative data for longitudinal studies of larger populations of students.

Until now we have mainly tested Talebob as a didactic tool to enhance the spoken language teaching in a classroom setting. However, as we do log all user productions and shall continue doing so for new versions, Talebob is not only useful as a didactic tool, but as a generator of substantial amounts of experimental data of a linguistic data type that can otherwise be difficult to elicit, exhibiting the pronunciation patterns of L2 learners in great detail. To our knowledge, no-one has produced a quantitatively based comparative study of the pronunciation patterns of Danish students. We are currently compiling data for such a paper, charting the pronunciation habits (and skills) as a function of their own first language, their prior exposure to Danish, their gender and age, self-declared degree of motivation, etc.

We thus wish to point to Talebob as an example of CALL-based screening of large groups of pupils. Access to statistical information about the progress of individual pupils, classes, and even populations of classes may be useful even for political decision-makers. Such considerations are highly relevant in Denmark right now, the 2014 school reform currently being implemented. For the first time ever English is now taught from first grade in Denmark. Spokesmen for the teachers are constantly expressing concerns about the lack of training programmes for teachers new to the challenge of teaching English to minors. Objective means for assessing the learning patterns are frequently called for in the press and in parliament. We believe that cleverly designed CALL-tools could play a decisive role in this debate.

8 A concluding remark

After having tested Talebob extensively for almost six months now with L2 learners of Danish in three countries, our most significant overall opservation is that pupil users generally *like* Talebob and spend far more time (at home and in school) training Danish pronunciation than ever before (Hauksdottir and Henrichsen (in prep.)). We have not yet performed any quantitative evaluations of the didactic effects, so we do not know whether Talebob can actually teach pupils a better Danish. Nevertheless, teachers in our test group (especially Icelanders) report that most of their pupils never practiced spoken Danish before unless forced. A majority of pupils report that they feel more confident now when using Danish speech productively (Hauksdottir and Henrichsen (in prep.)). This result alone, we feel, have made Talebob a worthwhile enterprise.

Acknowledgments

The presented work is a part of the ongoing Nordic project "Talehjælp til Dansk som Nabosprog" 2013-2015, supported by NorFA and Nordisk Ministerråd/Nordplus. We gratefully acknowledge their contributions. The project combines didactic and computational-linguistic research in Iceland, Denmark, and Sweden with practical implementation work by language teachers in Nuuk, Hafnarfjörður and Tórshavn (visit http://www.taleboblen.hi.is). Many have thus contributed, from a geographical area spanning five time zones. One, however, outshines all others: project leader and initiator Auður Hauksdóttir. Thanks to Auður for her many years as a powerstation in Nordic L2 didactics.

References

de Vries, B. P., Cucchiarini, C., Bodnar, S.. 2014. *Automatic Feedback on Spoken Dutch of Low-Educated Learners: An ASR-based CALL study*. Proceed. of EUROCALL 2014 (to appear).

Mbah, E.E., Mhab, B.M., Iloene, M.I., Iloene, G.O. 2013. *Podcasts for Learning English Pronunciation in Igboland: Students' Experiences and Expectations*. EUROCALL 2013.

González, J.F. 2012. *Can Apple's iPhone Help to Improve English Pronunciation Autonomously? State of the App*. EUROCALL 2012

Appel, C., Robbins, J., Moré, J., Mullen, T. 2012. *Task and Tool Interface Design for L2 Speaking Interaction Online*. EUROCALL 2012

Hauksdottir, A., Henrichsen, P.J. (in prep.) *Danskfaget i Vestnorden og det Digitale Læremiddel Taleboblen*

Thorborg, L. 2006. *Dansk Udtale i 49 Tekster*. Synope, ISBN 87-91909-01-5 (cd and book)

Thorborg, L. 2003. *Dansk Udtale - Øvebog*. Synope, ISBN 87-988509-4-6 (cd and book)

Selsøe Sørensen, H., Henrichsen, P.J., Hansen, C. 2004. *NorFA CALL net: Nordisk Netværk om Computerstøttet Unvervisning i Nordiske Sprog*; Nordisk Sprogteknologisk Forskningsprogram 2000-2004, Samfundslitteratur Press, 224pp.

Wang, H., T. Kawahara and Y. Wang. 2011. *Improving Non-native Speech Recognition Performance by Discriminative Training for Language Model in a CALL System;* INTERSPEECH 2011, 27-31

Mirzaei (2014) *Partial and synchronized captioning: A new tool for second language listening development;* EUROCALL 2014.

Henrichsen, P.J. 2004. *The Twisted Tongue; Tools for Teaching Danish Pronunciation Using a Synthetic Voice*; in Henrichsen (2004b)

Henrichsen, P.J. 2004b. *"CALL for the Nordic Languages - tools and methods for Computer Assisted Language Learning;* Cph. Studies in Language 30/2004

Allwood, J. and Henrichsen, P.J. (eds). 2005. *SweDanes for CALL - A corpus and computer-based student's aid for comparison of Swedish and Danish spoken language.* NorFA CALL NET (cd with manual).