DEALING WITH CROSS-LINGUAL ASPECTS IN SPOKEN NAME RECOGNITION

Frederik Stouten[1], *Jean-Pierre Martens*[1]

[1]ELIS, University of Ghent, Ghent, Belgium

fstouten@elis.ugent.be, martens@elis.ugent.be

ABSTRACT

The development of an automatic speech recognizer (ASR) that can accurately recognize spoken names belonging to a large lexicon, is still a big challenge. One of the bottlenecks is that many names contain elements of a foreign language origin, and native speakers can adopt very different pronunciations of these elements, ranging from completely *nativized* to completely *foreignized* pronunciations. In this paper we further develop a recently proposed method for improving the recognition of foreign proper names spoken by native speakers. The main idea is to combine the standard acoustic model scores with scores emerging from a phonologically inspired back-off model that was trained on native speech only. This means that the proposed method does not require the development of any foreign phoneme models on foreign speech data. By applying our method on a baseline Dutch recognizer (comprising Dutch acoustic models) we could reduce the name error rate for French and English names by a considerable amount.

Index Terms: spoken name recognition, phonological features, cross-lingualism

1. INTRODUCTION

It is a challenge to develop an automatic speech recognizer (ASR) that can accurately recognize proper names (e.g. person names, city names, street names, etc.) when the perplexity of the task is elevated. In a directory assistance application for instance, there may be a few 100K person names to distinguish, and it would be extremely expensive to elicit from human experts typical phonetic transcriptions of each name. Hence, one must rely on an automatic grapheme-to-phoneme (G2P) converter instead. But unfortunately commercially available G2P converters were trained to transcribe the regular words of a language. When confronted with foreign names, they often do not produce an acceptable output. Recent experiments on the transcription of person and geographical names occurring in the Netherlands showed that the state-of-the art Dutch G2P converter of Nuance was unable to produce an acceptable phoneme sequence (one of the manual transcriptions present in a lexical database) for about 30% of these names. When also considering wrong lexical stress assignments as errors, the error rate further increased to 50% [1].

Even if the G2P converter could be improved to produce more often an acceptable transcription, there would still be a problem because there is clear evidence (e.g. [2]) that, depending on their familiarity with the language of origin, native speakers may use different pronunciations of a foreign name. These pronunciations can range from totally *nativized* pronunciations (using native phonemes and native G2P rules) to totally *foreignized* pronunciations (using foreign phonemes and foreign G2P rules). We therefore argue that the ASR should incorporate lexical and acoustic models that can cope with this type of pronunciation variability.

In [3] one proposes to use multiple G2Ps to produce multiple pronunciations of a name: one G2P for the native language and one for each likely language of origin of the name. These likely languages of origin are determined by running a language identification algorithm on the name. Obviously, the outputs of the non-native G2Ps must be converted to native phoneme sequences that are compatible with the acoustic models of the native ASR. Adding the obtained pronunciations to the baseline dictionary caused a reduction of the word error rate (WER) by 25% for foreign names spoken by foreign speakers and by 10 % for foreign names spoken by native speakers, the case we will be dealing with in the present paper.

In [4], one also creates pronunciation variants, but this time in a data-driven way. This is achieved by using native acoustic models to align each name utterance with a graph of available initial pronunciations of that name (6 per name) as identified on the basis of expert knowledge. By seeking alternative phonemes for modeling the regions where the acoustics badly match the graph, new pronunciations were created. Including these pronunciations in the lexicon resulted in an improvement of the name recognition error rate by 20 to 40% relative. However, these figures may be optimistic because the tests were run on the same names that were also used to learn the new pronunciations.

A number of authors argue that in order to perform well, some non-native phonemes should be kept in the phonetic transcriptions and separate acoustic models should be available for these phonemes. In [5] for instance, models of English phonemes that have no good German equivalent were trained on English speech spoken by German speakers and added to the inventory of acoustic models. By doing so the WER on a corpus of German sentences containing at least one English name dropped from 60 to 44%.

In [6], non-native pronunciation variants for names of an English origin are generated in a totally data-driven way. An English phoneme recognizer generates English pronunciations, and by aligning these pronunciations with the canonical pronunciations emerging from a German G2P converter, one obtains training examples for the automatic learning of decision trees that can be used for the generation of English-accented pronunciation variants. This method however only yields a small drop (5.2 % relative) of the WER .

In cases where names from several languages have to be recognized, an approach that needs foreign phoneme models for each of these languages may turn out to be impractical, especially when some of these languages are less-resourced languages such as Indonesian, Russian, etc.. In that case one can try to create acoustic models for all the sounds in the IPA (International Phonetic Alphabet) and use these models for the mapping of foreign phonemes to symbols that have an associated acoustic model (e.g. [7]).

In a recent paper [8] we proposed a novel method that is a bit related to the just mentioned IPA approach, in the sense that it uses a phonologically motivated back-off score in combination with the traditional acoustic likelihoods. Preliminary experiments showed that the method can be effective even with a back-off model that was exclusively trained on native speech data. This confirms an earlier finding [9] that phonological feature models learned on native speech are also capable of characterizing foreign sounds.

In this paper we further elaborate and motivate our model and we assess its capabilities on a substantial trilingual spoken name corpus. In a later stage, we also plan to investigate the capabilities of our method in combination with phonological models that were trained on multilingual speech data, because it was shown in [10] that such models are more reliable than monolingually trained models.

The outline of this paper is as follows. In section 2 we recall the basics of our previously proposed method, but we present better motivations and a stepwise introduction of the foreignizable phoneme concept. The main contribution of the paper resides in the presentation, in Section 3, of a new and much more extensive experimental study than the one presented in the original paper. The major conclusions of this study and two directions for future research can be found in Section 4.

2. METHODOLOGY

Suppose that q represents a state of a baseline acoustic model, and that $\log p_A(\mathbf{x}|q)$ is the log-likelihood of acoustic vector **x** in this state. Suppose further that $\log p_B(\mathbf{x}|q)$ is the log-likelihood of acoustic vector **x** in state q as computed on the basis of a phonologically inspired back-off model. In that case, we propose to replace the standard acoustic score by a two-stream score

$$LL(\mathbf{x}|q) = g_{1q} \log p_A(\mathbf{x}|q) + g_{2q} \left[\alpha \ \log p_B(\mathbf{x}|q) - \beta\right] \quad (1)$$

with g_{1q} and g_{2q} being **state dependent** stream weights, and (α, β) normalization coefficients whose role will be explained in a moment.

2.1. Computing the phonological score

The core of the back-off model is a neural network that computes the posterior probabilities $P(f_i|\mathbf{x})$ of the 25 binary phonological features (PHFs) f_i $(i = 1, .., 25)$ that form the basis for our phonological description of acoustic model states [11]. The binary features describe (1) the **vocal source** (voiced, inactive), (2) the **manner of articulation** (closure, vowel, fricative, burst, nasal, approximant, lateral, silence), (3) the **place of articulation of consonants** (labial, labio-dental, dental, alveolar, post-alveolar, velar, glottal) and (4) the **articulatory properties of vowels** (low, mid-low, mid-high, high, back, mid, front, retroflex, rounded).

Since the PHF detector computes posterior probabilities, the log-likelihoods can be computed as

$$\log p_B(\mathbf{x}|q) = \log \frac{P_B(q|\mathbf{x})}{P_B(q)} + \log p(\mathbf{x}) \quad (2)$$

If we further assume that $g_{2q} \log p(\mathbf{x})$ is only weakly dependent on the state q, it does not contribute much to the discrimination between states and it can therefore be ignored. In that case, we argue that

$$LL(\mathbf{x}|q) = g_{1q} \log p_A(\mathbf{x}|q) + g_{2q}[\alpha \ \log \frac{P_B(q|\mathbf{x})}{P_B(q)} - \beta] \quad (3)$$

would be an acceptable two-stream score to use. Now it is time to explain what the role of (α, β) is. We first aligned the training data with the baseline models so that each frame was assigned a state q. Then α and β was chosen such that

$$\{\text{E,Var}\}[\alpha \log \frac{P_B(q|\mathbf{x})}{P_B(q)} - \beta] = \{\text{E,Var}\}[\log p_A(\mathbf{x}|q)] \quad (4)$$

taken over all frames. This makes the two stream scores more equivalent, and (g_{1q}, g_{2q}) interpretable as stream importances. The search for the optimal stream weights can then be restricted to $g_{1q} + g_{2q} = 1$.

To compute the back-off score from the posterior probabilities $P(f_i|\mathbf{x})$, we need a PHF characterization of state q. This characterization is derived from the PHF characterization of the corresponding phoneme and from the average $P(f_i|\mathbf{x})$ of observations assigned to state q in a forced alignment (see [8]). If P_q denotes the set of *positive* features that are supposed to be *on* for this state, and N_q the complementary set of negative features, then, assuming independent PHFs leads to

$$\log \frac{P_B(q|\mathbf{x})}{P_B(q)} = \sum_{f_i \in P_q} \log \frac{P(f_i|\mathbf{x})}{P(f_i)} + \sum_{f_i \in N_q} \log \frac{1-P(f_i|\mathbf{x})}{1-P(f_i)} \quad (5)$$

Now it happens that the prior probabilities of the features are usually low (because they are only positive for some of the phonemes) and that the posterior probabilities for the negative features in the correct state are even lower. This means that the ratios corresponding to the negative features are usually close to one on the states of the optimal state sequence, and consequently, that they do not contribute much to the phonological score. However, if this score is dominated by the positive features, we have to take into account that different states have a different number of positive features. This means that there is a danger that the contributions of the phonological scores to the two-stream score on different states are not compatible. To compensate for this, we have replaced the sums by means in Equation 5:

$$\log \frac{P_B(q|\mathbf{x})}{P_B(q)} = \frac{1}{\text{card}(P_q)} \sum_{f_i \in P_q} \log \frac{P(f_i|\mathbf{x})}{P(f_i)} + \frac{1}{\text{card}(N_q)} \sum_{f_i \in N_q} \log \frac{1-P(f_i|\mathbf{x})}{1-P(f_i)} \quad (6)$$

It has been verified experimentally [8] that this replacement gives an improvement.

2.2. Determination of the stream weights

In order to determine optimal stream weights for each state q, we would have to conceive an automatic weight optimization scheme. However, before starting to develop such a scheme, we will investigate what can be achieved with state-independent stream weights (g_1, g_2) which are optimized by tracking the WER, measured on a development set, as a function of $g_2 = 1 - g_1$, and by selecting the value yielding the minimal WER.

2.3. Phonological characterization of a state

Until now, the phonological description of state q emerges from PHF characterization of the phoneme from which it originates. However, we argue that it makes sense to make a distinction between two types of phonemes: (i) purely native phonemes and phonemes originating from foreign phonemes with the same phonological description on the one hand, and (ii) phonemes originating from a foreign phoneme with a deviating phonological description on the other hand. We call the latter phonemes *foreignizable* and we anticipate that they are often pronounced with the phonological features of their foreign counterparts. Therefore, we use the phonological representation of the foreign phoneme in that case to compute the back-off score for all the states of that phoneme.

In order to do so, foreignizable phonemes are explicitly marked in the lexicon by means of an underscore notation (see Table 1). When /r_rr/ appears in an English name that is part of a Dutch lexicon, it means that the Dutch phoneme /r/ (from the Dutch word *oo***r**) was obtained as an approximation of the English /rr/ (from the English word *o***r**) and that the acoustic score must be obtained by combining the model score emerging from a triphone model with /r/ as the central phoneme and a back-off score computed on the basis of the PHFs of /rr/. Note that it can happen (see Table 1) that two subsequent phonemes (e.g. the Dutch /Y/ (from *b***u***s*) + /r/) originate from the same foreign phoneme (e.g. the English /3:/ from *b***i***rd*) and vice versa.

Table 1. *Two English names with their baseline and alternative transcriptions comprising foreignizable phonemes (symbols are SAMPA, except for the English /r/ which is denotes as /rr/). A hyphen represents a short pause.*

name	transcription	
Burr Tuppel	baseline	b Y r - t Y p @ l
	alternative	b Y_3: r_3: - t Y p @ l
Alan Presser	baseline	E l @ n - p r E s @ r
	alternative 1	E l @ n - p r_rr E s @ r
	alternative 2	E l @ n - p r E s @ r_rr
	alternative 3	E l @ n - p r_rr E s @ r_rr

The number of foreignizable phonemes depends on the (native, foreign) language combination: for (Dutch, English) we found 6 foreignizable phonemes (see [8]), for (Dutch, French) we found 7.

2.4. Introduction of pronunciation variants

Foreignizable phonemes can also form a basis for the generation of pronunciation variants in the lexicon. A simple way to accomplish this is to produce alternative pronunciations by replacing one or more foreignizable phonemes by their pure native equivalents. Table 1 shows two names and the variants that were created for them in this way. The underlying motivation is that the user may adopt a nativized pronunciation for all or just for some of the foreignizable phonemes. In that case it may be advantageous to let the recognizer decide where to select nativized and where to select foreignized pronunciations.

3. EXPERIMENTS

The experiments in [8] were restricted to the recognition of English names by a Dutch speech recognizer, and the number of different English names was quite limited. In this paper we perform tests on a much larger corpus of spoken names, and report results for English, French as well as Dutch names, uttered by Dutch speakers.

The spoken name corpus was recorded in the AUTONOMATA project that was funded by the Dutch-Flemish STEVIN program [12]. The database will soon be made publicly available by Dutch-Flemish Language & Speech Technology Center (www.tst.inl.nl). In the present study we selected the 60 Dutch speakers from Flanders (one of the two regions in Europe were Dutch is spoken). Each speaker uttered one of 10 lists of 120 Dutch, 23 English, 23 French and 15 Moroccan names and there was no overlap between these 10 name lists. One third of the speakers was between 12 and 18 years old, the remaining speakers were adults. The names were either person names (first name + family name), city names or street names.

In the present study the Moroccan names were omitted and the remaining data were divided in an adaptation set, a development set and a test set (see Table 2). The same 60 speakers were present in all the data sets, but the adaptation set comprised only Dutch names, and there was no overlap in the names occurring in the development set and the test set. The test set was designed to contain a large percentage of foreign names.

Table 2. *Composition (in terms of language of origin of the name) of the test, development and adaptation sets extracted from the AUTONOMATA spoken name corpus.*

	English (E)	French (F)	Dutch (D)	All
adapt	–	–	4440	4440
develop	380	380	760	1520
test	1000	1000	2000	4000
total	1380	1380	7200	9960

In all experiments the ASR had a vocabulary of 1660 names: 1200 Dutch, 230 English and 230 French names. However, the ASR is assumed to have no prior knowledge of the language of origin of these names. The acoustic models were triphone models: either speaker-independent models (SIMs) that were trained using HTK [13] on a multi-speaker read speech corpus recorded in the Flanders [14], or adapted models (AMs) obtained from these SIMs by MLLR adaptation on the basis of the adaptation set extracted from the spoken name corpus. During adaptation we trained 32 model transformation matrices according to the procedure explained in the HTK-book.

Although the adaptation set contains the same speakers as the test set, it was verified in a separate experiment with a smaller test set and no speaker overlap between the adaptation set and the test set, that the WERs on the test set were very similar to those reported here. This means that the models are not so much adapting to the test speakers, but mainly to the acoustic circumstances appearing in the spoken name corpus recordings.

We will now describe the baseline experiments that have been run, and after that, the experiments that were conducted to assess the capabilities of our method.

3.1. Setting up a baseline system

In the baseline system, no back-off models nor pronunciation variants created based on foreignizable phonemes were used.

We have investigated the effect of using different types of transcriptions in the lexicon. To that end we had available the Dutch, English and French versions of the Nuance G2P-converter, and an example transcription of each name. The latter is a transcription that, according to a human expert, is a likely and acceptable transcription of the name. Using these resources we composed the following lexicons:

DuAlone	all names transcribed by Dutch G2P
All	all names transcribed by three G2Ps
ManAlone	manual transcriptions of all names
DuMan	merge of DuAlone and ManAlone
AllMan	merge of All and ManAlone

The corresponding word (name) error rates obtained with the two acoustic model sets on the different parts of the test set are listed in Tables 3 (SIMs) and 4 (AMs).

Table 3. *Baseline performances (WER in %) obtained with the speaker independent acoustic models in combination with the different lexicons.*

lexicon	English	French	Dutch	All
DuAlone	61.7	43.3	19.3	35.9
All	50.8	32.5	21.3	31.5
ManAlone	45.1	47.5	17.5	31.9
DuMan	42.7	37.4	17.7	28.9
AllMan	47.0	33.5	19.8	30.0

Table 4. *Baseline performances (WER in %) obtained with the adapted acoustic models in combination with the different lexicons.*

lexicon	English	French	Dutch	All
DuAlone	33.7	23.4	4.2	16.4
All	20.7	12.8	4.4	10.6
ManAlone	15.7	30.7	3.7	13.4
DuMan	13.3	17.8	3.7	9.6
AllMan	15.1	14.8	3.9	9.4

The most important finding is that foreign G2Ps produce much better transcriptions of foreign names than the native

G2P, even with the foreign phonemes being mapped to native phonemes. This can only mean that a lot of native speakers adopt foreign name pronunciations that are closer to foreignized than to nativized pronunciations.

A second important finding is that the foreign G2Ps (*All*) offer transcriptions that outperform (especially with AMs) the manual transcriptions (*ManAlone*).

A third finding is that the Dutch transcriptions are indispensable to get a good result: *DuMan* significantly out performs *ManAlone*.

A last finding is that the foreign G2Ps do not attribute much anymore if the Dutch and manual transcriptions are already in the lexicon. For the Dutch names they are useless and only augmenting the lexical confusion, whereas for foreign names there is a balance between that effect and the positive effect of bringing in a better transcription than the manual one for some of these names.

3.2. Testing the proposed methodology

Since one usually has no access to manual transcriptions we take *All* as the baseline lexicon and we assess our methodology when applied in combination with this lexicon. Figures in bold in the Tables refer to results that are significantly better than the baseline according to a Wilcoxon signed-rank test [15] with $p = 0.05$.

3.2.1. Back-off model with native phoneme representations

In a first experiment we just took the lexicon *All* as used in the baseline system. The phonological representations that served as a basis for the computation of the back-off scores were those of the native phonemes that gave rise to the model states. We determined the optimal stream weight by performing recognition tests on the development set for several values of g_2. We tracked the WER as a function of g_2, smoothed the curve and located the minimum of the smoothed curve. The corresponding stream weights were then imputed in the ASR system. The optimal stream weights were (g_1, g_2) = (0.2,0.8). The corresponding WERs plus the absolute and relative improvements (AI and RI) over the baseline are summarized in Table 5. The first remarkable fact is that the improvement is modest in the SIM case but substantial in the AM case. Possibly, our method is not effective as long as the baseline acoustic models are insufficiently accurate.

A second remarkable fact is that in the AM case, the improvement is not only substantial for English and French names, but surprisingly, also for Dutch names. Apparently, the back-off model provides information that is not captured by the triphone model.

3.2.2. Back-off model with foreign phoneme representations

In a second experiment, we replaced the former *All* lexicon by a lexicon with foreignizable phonemes in the foreign G2P outputs of foreign names. Then we used the phonological characterization of the foreign phonemes to control the back-off score computation. The results of this experiment are summarized in Table 6. Obviously, the introduction of foreign phonological representations causes only a small consistent gain, but one that is achievable at no extra cost.

Table 5. *Performances (all in %) of an ASR with a two-stream acoustic model and native phonological representations of the model states.*

triphones	measure	E	F	D	All
SIMs	WER	**47.3**	30.8	21.0	**30.0**
	AI	**3.5**	1.7	0.2	**1.5**
	RI	**6.9**	5.2	1.0	**5.5**
AMs	WER	**18.3**	**10.2**	**3.1**	**8.7**
	AI	**2.4**	**2.6**	**1.3**	**1.9**
	RI	**11.6**	**20.3**	**29.5**	**17.9**

Table 6. *Performances (all in %) of an ASR with a two-stream acoustic model and foreignizable phonological representations of the model states.*

triphones	measure	E	F	D	All
SIMs	WER	**46.7**	30.1	21.2	**29.8**
	AI	**4.1**	2.4	0.1	**1.8**
	RI	**8.0**	7.4	0.7	**5.5**
AMs	WER	**18.1**	**10.1**	**3.1**	**8.6**
	AI	**2.6**	**2.7**	**1.3**	**2.0**
	RI	**12.6**	**21.1**	**29.5**	**18.9**

One of the possible explanations for the low gain is that the speakers not always use a foreign pronunciation, and thus that a back-off model on the basis of a foreign representation is not always the best solution. In order to test that hypothesis we have conducted an additional experiment.

3.2.3. Including pronunciation variants

In a third experiment we have introduced pronunciation variants in the lexicon using the method proposed in section 2.4. The recognition results obtained with this lexicon are summarized in Table 7.

The Table reveals that the results have further improved, and that the improvement is now starting to be statistically significant for the speaker-independent case as well. Note too, that the improvement is confined to the English and French name subsets, as expected. However, the gain is moderate and adding variants is increasing the computational load. We

Table 7. *Performances (all in %) of an ASR with a two-stream acoustic model,foreignizable phonological representations of the model states and pronunciation variants.*

triphones	measure	E	F	D	All
SIMs	WER	**45.9**	**29.4**	21.4	**29.5**
	AI	**4.9**	**3.1**	-0.1	**2.0**
	RI	**9.6**	**9.5**	-0.0	**6.3**
AMs	WER	**17.6**	**10.0**	**3.1**	**8.5**
	AI	**3.1**	**2.8**	**1.3**	**2.1**
	RI	**14.9**	**21.9**	**29.5**	**19.8**

therefore we recommend to use the system with foreignizable representations but without variants.

4. CONCLUSIONS AND FUTURE WORK

We have further elaborated a novel technique for improving the recognition of foreign names spoken by native speakers. The method is based on the introduction of a two-stream acoustic model and foreignizable phonemes in the lexicon. The two-stream acoustic models combine the standard acoustic likelihood on a triphone state with a phonological score for that same state. The standard acoustic models and the phonological feature extractors were both trained on native speakers only.

For the recognition of English and French names spoken by Dutch speakers, the method yielded significant reductions of the WER of 15 % and 22 % relative. Surprisingly, the recognition also improved for the Dutch names.

We are currently evaluating our methodology under the assumption that the ASR knows the language of its vocabulary entries. Furthermore, there is evidence [8] that the results can be further improved by training the phonological feature extractor on multilingual speech.

5. ACKNOWLEDGMENTS

This work was supported by the Flemish Institute for the Promotion of Scientific and Technical Research in the Industry (contract SBO/40102). Nuance is acknowledged for making available the G2P converters.

6. REFERENCES

[1] H. van den Heuvel, J.-P. Martens, and N. Konings, “G2p-conversion of names. what can we do (better)?,” in *Proc. Interspeech*, Antwerp, Belgium, 2007, pp. 1773–1776.

[2] S. Fitt, “The pronunciation of unfamiliar native and non-native town names,” in *Proc. Eurospeech*, Madrid, Spain, 1995, pp. 2227–2230.

[3] B. Maison, S.F. Chen, and P.S. Cohen, “Pronunciation modeling for names of foreign origin,” in *Proc. ASRU*, Virgin Islands, USA, 2003, pp. 429–434.

[4] F. Beaufays, A. Sankar, S. Williams, and M. Weintraub, “Learning name pronunciations in automatic speech recognition systems,” in *International Conference on Tools with Artificial Intelligence*, Washington, USA, 2003, pp. 233–240.

[5] G. Stemmer, E. Nöth, and H. Niemann, “Acoustic modeling of foreign words in a german speech recognition system,” in *Proc. Eurospeech*, Aalborg, Denmark, 2001, pp. 2745–2748.

[6] S. Goronzy, S. Rapp, and R. Kompe, “Generating non-native pronunciation variants for lexicon adaptation,” in *Speech Communication*, 2004, vol. 42, pp. 109–123.

[7] S. Kunzmann, V. Fisher, J. Gonzalez, C. Emam, C. Günther, and E. Janke, “Multilingual acoustic models for speech recognition and synthesis,” in *Proc. ICASSP*, Montreal, Canada, 2004, pp. 745–748.

[8] F. Stouten and J.-P. Martens, “Recognition of foreign names spoken by native speakers,” in *Proc. Interspeech*, Antwerp, Belgium, 2007, pp. 2133–2136.

[9] G. Williams, M. Terry, and J. Kaye, “Phonological elements as a basis for language-independent asr,” in *Proc. ICSLP*, Sydney, Australia, 1998, pp. 88–91.

[10] S. Stüker, T. Schultz, F. Metze, and A. Waibel, “Multilingual articulatory features,” in *Proc. ICASSP*, Hong Kong, China, 2003, pp. 144–147.

[11] F Stouten and J.-P. Martens, “On the use of phonological features for pronunciation scoring,” in *Proc. ICASSP*, Toulouse, France, May 2006, pp. 229–232.

[12] E. D’Halleweyn, J. Odijk, L. Teunissen, and C. Cucchiarini, “The dutch-flemish hlt programme stevin: Essential speech and language technology resources,” in *Proc. LREC*, 2006, pp. 761–766.

[13] S. Young, D. Kershaw, J. Odell, D. Ollasson, V. Valtchev, and P. Woodland, “The htk-book version 3.0,” in *Cambridge University, Engineering Department*, 2000.

[14] K. Demuynck, D. Van Compernolle, C. Van Hove, and J.-P. Martens, “Een corpus gesproken nederlands voor spraaktechnologisch onderzoek,” in *Technical Report ESAT - ELIS*, 1997, pp. 1–30.

[15] W. Daniels, *Applied nonparametric statistics*, 1978.