# Appendix A

# Appendix

## A.1 Formant extraction using LPC - Matlab

### A.1.1 Autocorrelation

From Rabiner[1] we have:

If we have a windowed frame $s$ of size $N$ samples then the autocorrelation(with order $P$) is defined as:

$$R(i) = \sum_{n=0}^{N-1-i} s(n)s(n+i) \qquad i = 0, 1, ..., P. \tag{A.1}$$

In Matlab this can be coded as:

```
%       [R] = autocorr(s,P)
%       where s is the input vector, and P is the order of prediction.
%       Function to compute the autocorrelation of the data
%       computes autocorrelation R(i) for i=1, .. ,P+1.
```

```
function [R] = autocorr(s,P)

N=max(size(s));
for i=0:P
 R(i+1,1)=sum(s(1:N-i).*s(i+1:N));
end
```

### A.1.2 Durbin recursion

Durbin recursion (where we have $L$ frames) is defined in Rabiner[1] as:

$Solve\ recursively\ for\ i = 1, 2, ..., P:$

$$E(0) = R(0) \tag{A.2}$$

$$k_i = \frac{\{R(i) - \sum_{j=1}^{L-1} a_j^{(i-1)} R(\mid i-j \mid)\}}{E(i-1)} \qquad 1 \leq i \leq P \tag{A.3}$$

$$a_i(i) = k_i \tag{A.4}$$

$$a_j(i) = a_j(i-1) - k_i a_{i-j}(i-1) \tag{A.5}$$

$$E(i) = (1 - k_i^2)E(i-1) \tag{A.6}$$

In Matlab this can be coded as:

```
%  [a]=durbin(R);
%  Function to calculate the linear predictive coefficients a, from
%  autocorrelation lags R.

function [a] = durbin(R)
P = max(size(R))-1;
a = ones(P,1);
E(1)=R(1);
for i=1:P
 for j=1:i-1
  a_past(j)=a(j);
 end
 sum_term=0;
 for j=1:i-1
```

```
  sum_term = sum_term + a_past(j)*R(i-j+1);
 end
 k(i) = (R(i+1) - sum_term) / E(i);
 a(i) = k(i);
 for j=1:i-1
  a(j) = a_past(j) - k(i)*a_past(i-j);
 end

 E(i+1) = (1-k(i)^2)*E(i);
end
```

### A.1.3 Formant extraction

Utilising the functions above we can determine the formants for a frame of speech. The program simply utilises the LP coefficients (determined with the above functions) and a root finding algorithm to determine the resonance frequencies(formants) of the speech segment.

```
% [f] = formants(x,RO,NUM_FORMANTS,LPC_ORDER)
%   Function to estimate the NUM_FORMANTS formants of voiced speech x,
%   with LPC_ORDER order LPC analysis and peak picking. RO is a
%   parameter that varies between 0 and 1 and it is multiplied by each
%   LP coefficient to make the peaks clearer. It is usually 0.6.

function [f] = formants(x,ro,num_formants,LPC_ORDER,SAMP_FREQ);
 x=filter([1 -1],1,x);

 lpc=ro*durbin(autocorr(x,LPC_ORDER));
 f=roots([1 -lpc']);
 b=abs(SAMP_FREQ/2/pi*log10(abs(f)));
 f=SAMP_FREQ/2/pi*angle(f);
 f=f.*(f>200);
 index=find(f);
 f=f(index);
 b=b(index);
 [b,ind]=sort(b);
 f=f(ind);
 f=sort(f(1:num_formants));
end
```

## A.2 Pitch extraction using autocorrelation

- Step 1. Preprocessing: to remove the side-lobe of the Fourier transform of the Hanning window for signal components near the Nyquist frequency, a soft up-sampling is performed as follows: an FFT is performed on the whole signal; filtering is done by multiplication in the frequency domain linearly to zero from 95% of the Nyquist frequency to 100% of the Nyquist frequency; an inverse FFT of order one higher than the first FFT is then performed.

- Step 2. The global absolute peak value of the signal is computed(see Step 3.3).

- Step 3. Because the method is a short-term analysis method, the analysis is performed for a number of small segments (frames) that are taken from the signal in steps given by the TimeStep parameter (default is 0.01 seconds). For every frame at most MaximumNumberOfCandidatesPerFrame (default is 4) lag-height pairs are found that are good candidates for the periodicity of this frame. This number includes the unvoiced candidate, which is always present. The following steps are taken for each frame:

  Step 3.1. A segment is taken from the signal. The length of this segment (the window length) is determined by the MinimumPitch parameter, which stands for the lowest fundamental frequency that you want to detect. The window should be just long enough to contain three periods (for pitch detection) of MinimumPitch. E.g. if MinimumPitch is 75 Hz, the window length is 40 ms.

  Step 3.2. The local average is subtracted.

  Step 3.3. The first candidate is the unvoiced candidate, which is always present. The strength of this candidate is computed with two soft threshold parameters. E.g., if VoicingThreshold is 0.4 and SilenceThreshold is 0.05, this frame bears a good chance of being analysed as voiceless (in step 4) if there are no autocorrelation peaks above approximately 0.4 or if the local absolute peak value is less than approximately 0.05 times the global absolute peak value, which was computed in step 2.

Step 3.4. The segment is multiplied by a window function (e.g. Hanning).

Step 3.5. Half a window length of zeroes is appended (because autocorrelation values up to half a window length are needed).

Step 3.6. Zeroes are appended until the number of samples is a power of two.

Step 3.7. A Fast Fourier Transform is performed.

Step 3.8. The samples are squared in the frequency domain.

Step 3.9. A Fast Fourier Transform is performed. This gives a sampled version of $r_a(\tau)$.

Step 3.10. This is then divided by the autocorrelation of the window, which must be computed once with steps 3.5 through 3.9. This gives a sampled version of $r_x(\tau)$.

Step 3.11. The locations and heights of the maxima of the continuous version of $r_x(\tau)$ are then found. The only locations considered for the maxima are those that yield a pitch between MinimumPitch and MaximumPitch. The MaximumPitch parameter should be between MinimumPitch and the Nyquist frequency. The only candidates that are remembered, are the unvoiced candidate which has a local strength equal to

$$R \equiv VoicingThreshold + max\left(0.2 - \frac{\frac{(localabsolutepeak)}{(globalabsolutepeak)}}{\frac{(SilenceThreshold)}{(1+VoicingTheshold)}}\right) \qquad (A.7)$$

and the voiced candidates with the highest local strength

$$R \equiv r(\tau_{max}) - OctaveCost \cdot log_2(MinimumPitch \cdot \tau_{max}). \qquad (A.8)$$

The OctaveCost parameter favours higher fundamental frequencies. One of the reasons for the existence of this parameter is that for a perfectly periodic signal all the peaks are equally high and we should choose the one with the lowest lag. Another reason for this parameter is unwanted local downward octave jumps caused by additive noise.

After performing step 3 for every frame, a number of frequency-strength pairs $(F_{ni}, R_{ni})$ are left, where the index $n$ runs from 1 to the number of frames, and i is between 1 and the number of candidates in each frame. The locally best candidate in each frame is the one with the highest $R$. But as several approximately equally strong candidates can exist in any frame, a global path finder is utilised, the aim of which is to minimise the number of incidental voiced-unvoiced decisions and large frequency jumps.

- Step 4. For every frame $n$, $p_n$ is a number between 1 and the number of candidates for that frame. The values $p_n|1 \leq n \leq numberofframes$ define a path through the candidates: $(F_{np_n}, R_{np_n})|1 \leq n \leq numberofframes$. With every possible path a cost

$$cost(\{P_n\}) = \sum_{n=2}^{numberOfFrames} transitionCost(F_{n-1,p_{n-1}}, F_{np_n}) - \sum_{n-1}^{numberOfFrames} R_{np_n} \tag{A.9}$$

is associated, where the $transitionCost$ function is defined by

$$transitionCost(F1, F2) = \begin{cases} 0 & if\ F1\ unvoiced\ and\ F2\ unvoiced \\ VoicedUnvoicedCost & if\ F1\ unvoiced\ xor\ F2\ unvoiced \\ OctaveJumpCost \cdot |log_2 \frac{F1}{F2}| & if\ F1\ voiced\ and\ F2\ voiced \end{cases} \tag{A.10}$$

where the $VoicedUnvoicedCost$ and $OctaveJumpCost$ parameters could both be 0.2. The globally best path is the path with the lowest cost. This path might contain some candidates that are locally second-choice. The cheapest path can be found with the aid of dynamic programming, e.g., using the Viterbi algorithm described for Hidden Markov Models by Van Alphen and Van Bergem[44]. For stationary signals, the global path finder can easily remove all local octave errors, even if they comprise as many as 40% of all the locally best candidates. This is because the correct candidates will be almost as strong as the incorrectly chosen candidates. For most dynamically changing signals, the global path finder can still cope easily with 10% local octave errors.

## A.3 Pitch trajectories

### A.3.1 Vowel pitch trajectories

The figures in this section are the complete graphs of the pitch trajectories determined for the long vowels studied.

### A.3.2 Diphthong pitch trajectories

The figures in this section are the complete graphs of the pitch trajectories determined for the diphthongs studied.

## A.4 Expanded formant plots

### A.4.1 Expanded vowel formant plots

The graphs given in this section are the complete versions of the graphs shown in Figures 3.4 and 3.5. The individual utterance means are shown in addition to the

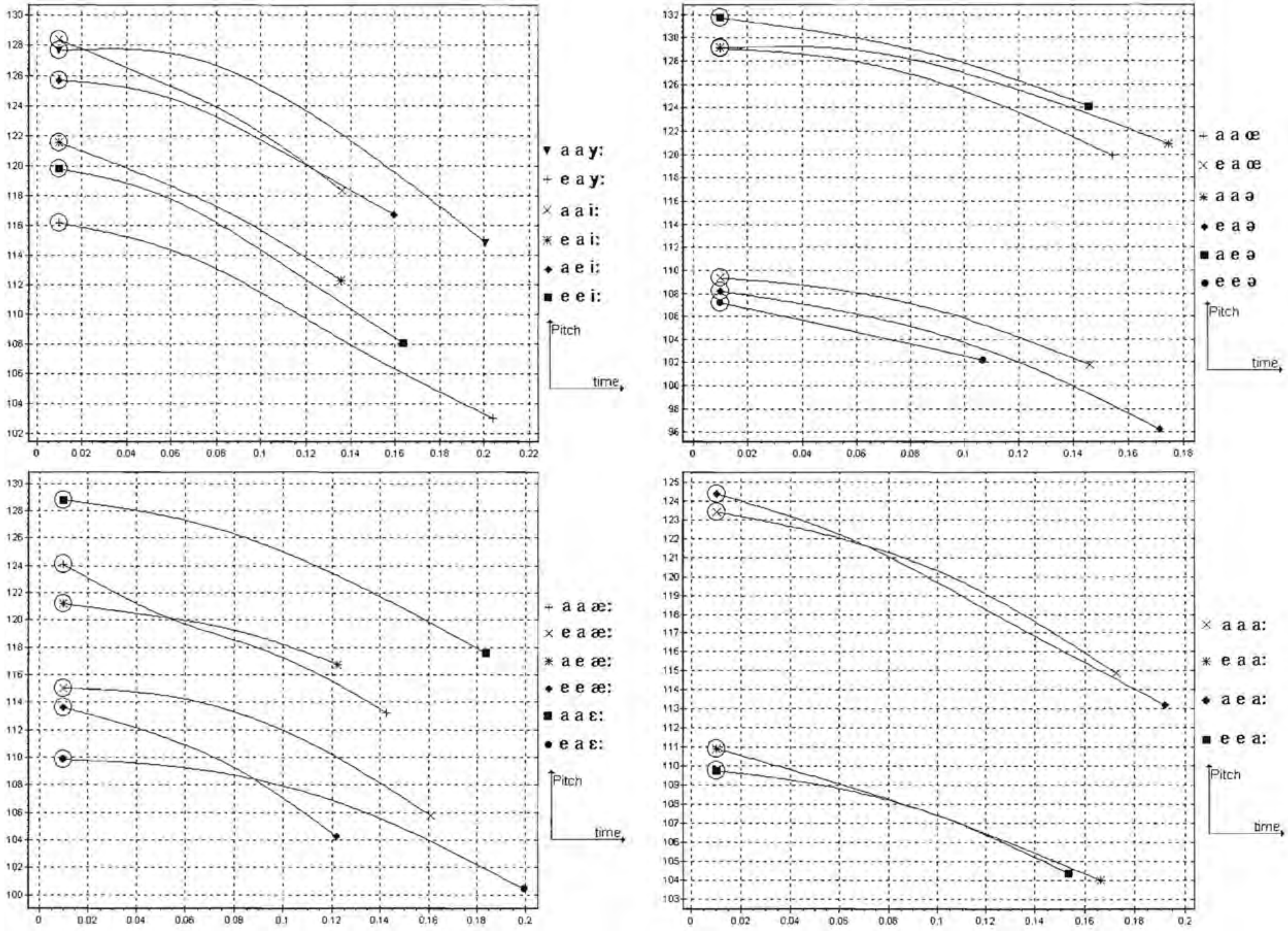


Figure A.1: Vowel pitch trajectories: [<y:> in "*uur*" and also <i:> in "*dier*" and "heat"], [<œ> in "*brûe*" and also <ə> in "*wîe*" and "about"], [<æ:> in "*werk*" and "hat" and also <ɛ:> in "*êrens*"] and [<a:> in "*klaar*" and "father"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from.



Appendix

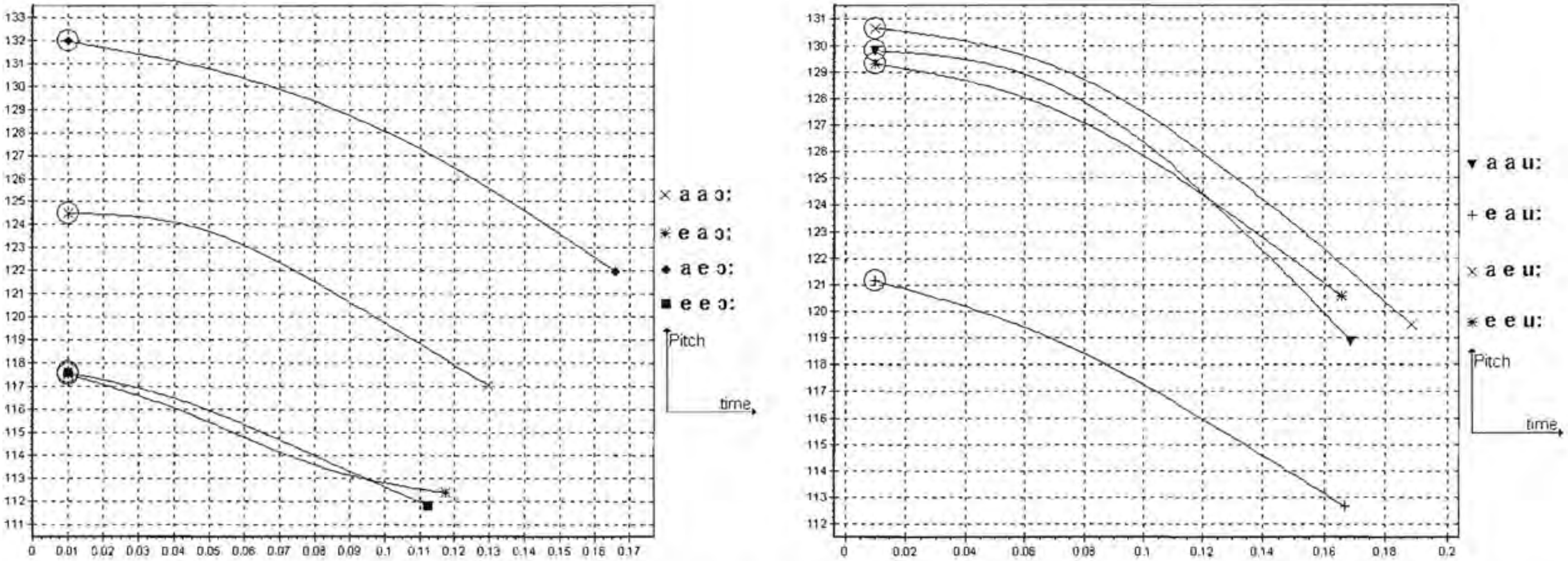


Figure A.2: Vowel pitch trajectories: [<ɔ:> in "*dom*" and in "bought"] and [<u:> in "*boer*" and "soon"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from.

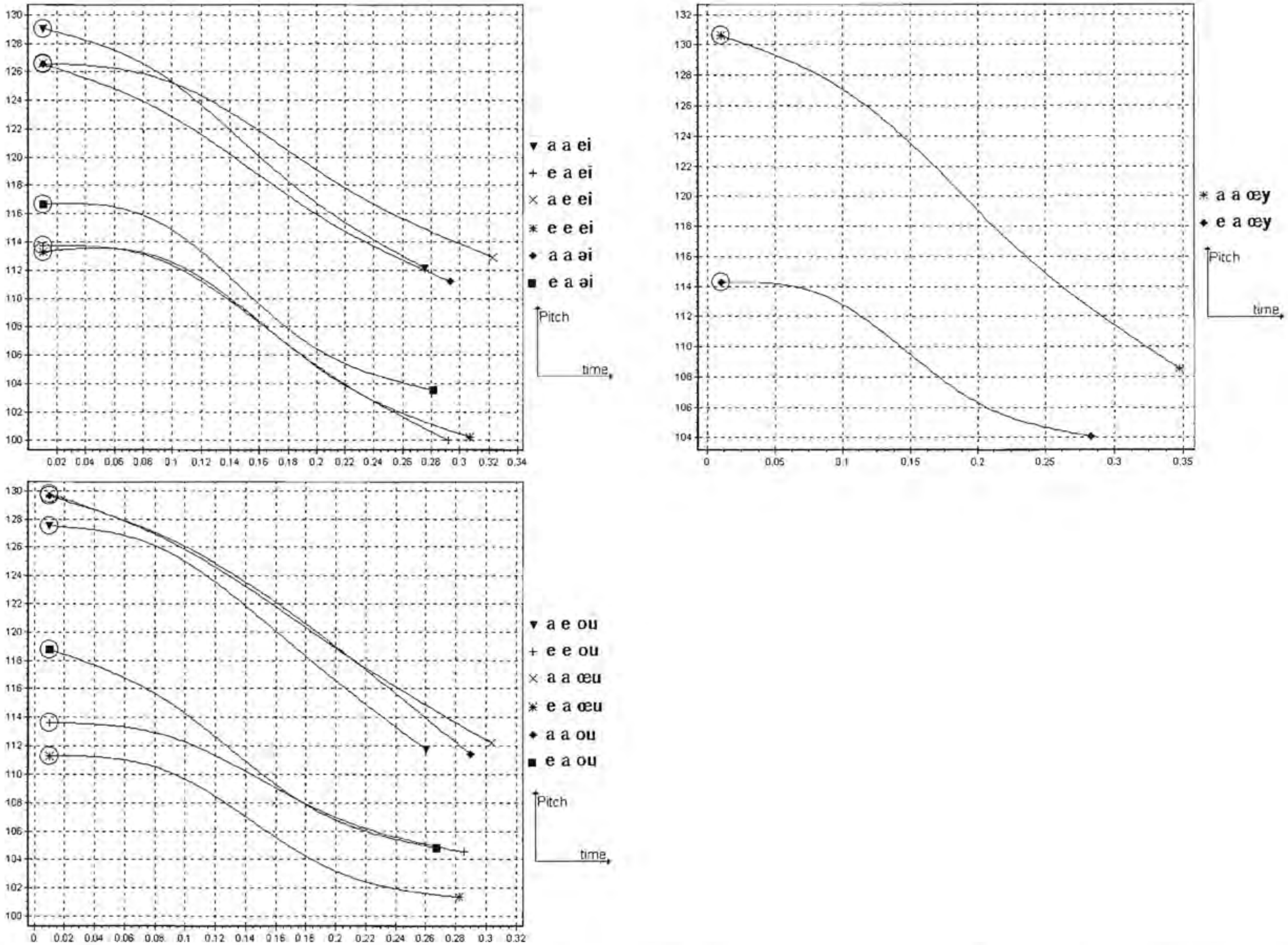


Figure A.3: Diphthong pitch trajectories: [<ei> in "*ryk*" and "play" and also <əi> in "*bly*"], [<œy> in "*trui*" ] and [<ou> in "*gou*" and "home" and also <œu> in "*blou*"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from.

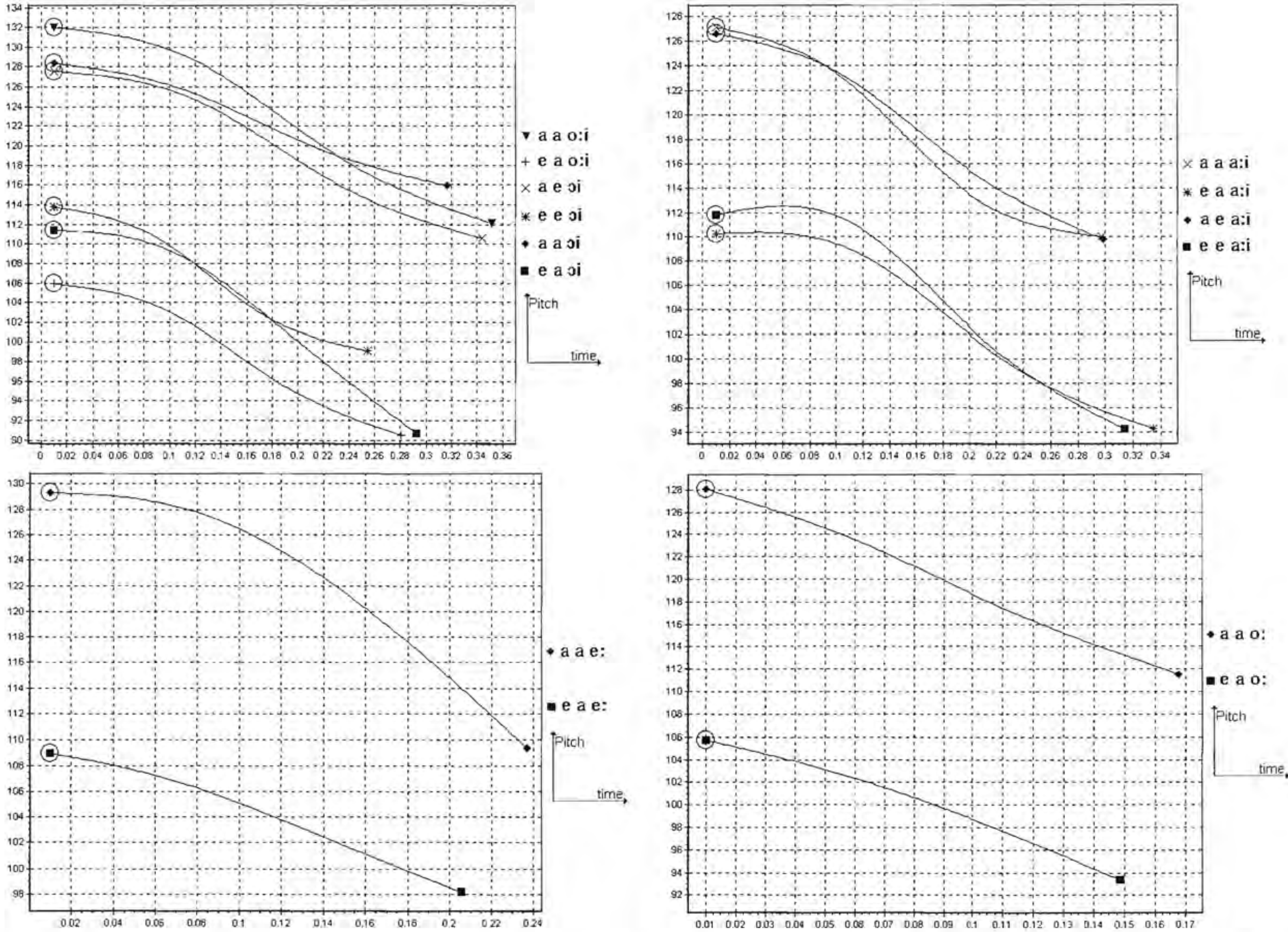


Figure A.4: Diphthong pitch trajectories: [<o:i> in "*mooi*" and also <ɔi> in "*hondjie*" and "boy"], [<a:i> in "*haai*" and "time"], [<e:> in "*bees*"] and [<o:> in "*kool*"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from.

global mean and variance as in the the simpler figures.

## A.5 Compact Disk Contents

The attached compact disk contains the following:

- The data recorded, labelled and used in the study.
- This dissertation in GZipped PostScript form.
- C Programmes

  Wyre: The programme used to segment and label the data.

  DataPlay: The programme used to play back the segmented sections for audio verification.

  DataSort: The programme used to split the data from speakers into language groups.

  Pitch: The programme used to convert Praat style pitch trajectory files into files suitable for GPlot.

  GPlot: The programme used to plot the mean vowel locations, variance bubbles, diphthong trajectories and perform analysis of variance comparisons.
- Matlab Programmes

  General: A number of programmes used to plot the results from research done in previous studies.

  SPTool: The programme used to verify that the extracted formants are correct when compared to the spectrograms.

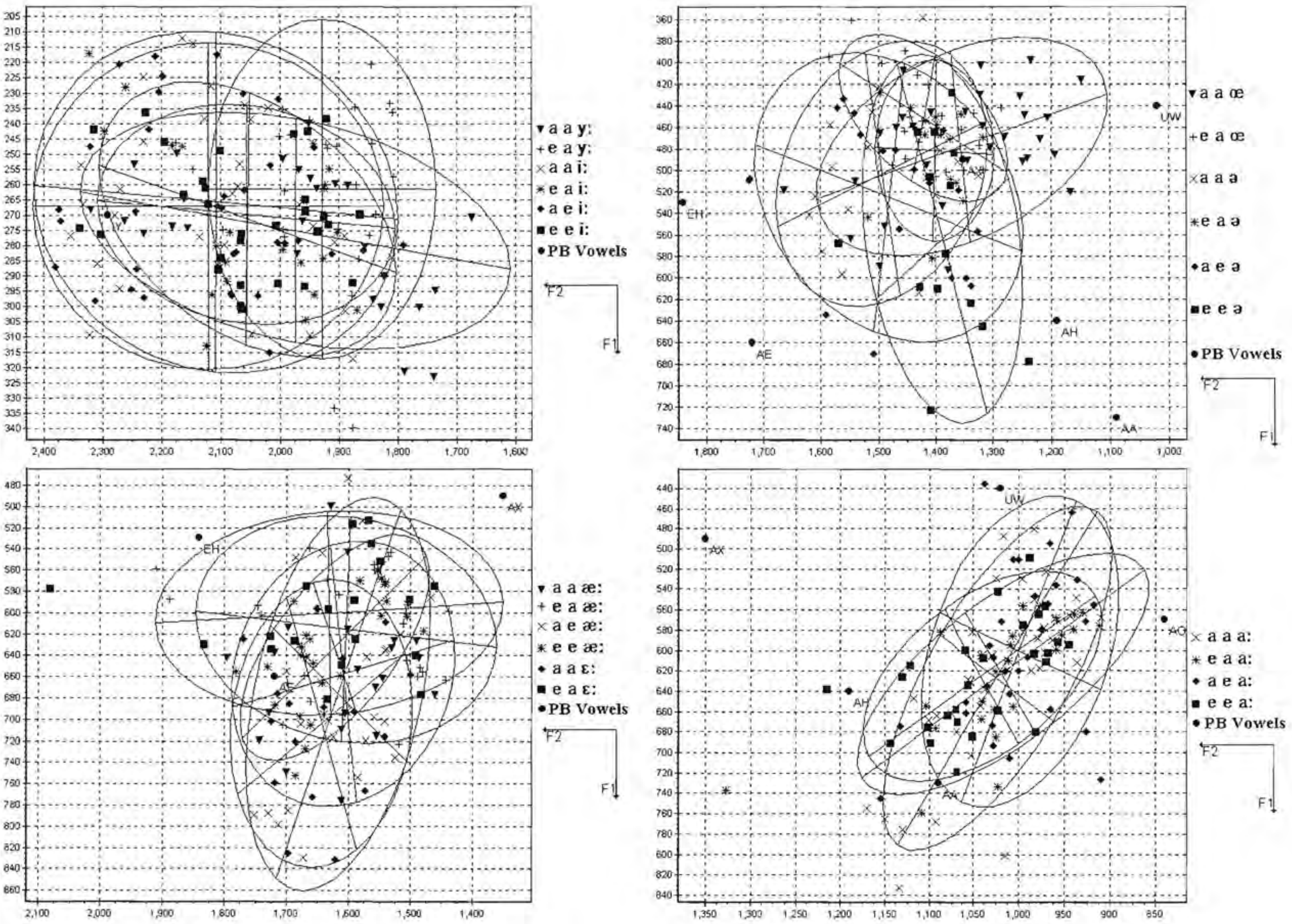


Figure A.5: Vowel formant clusters: [<y:> in "*uur*" and also <i:> in "*dier*" and "heat"], [<œ> in "*brûe*" and also <ə> in "*wîe*" and "about"], [<æ:> in "*werk*" and "hat" and also the incorrectly used <ɛ:> in "*êrens*"] and [<a:> in "*klaar*" and "father"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from.

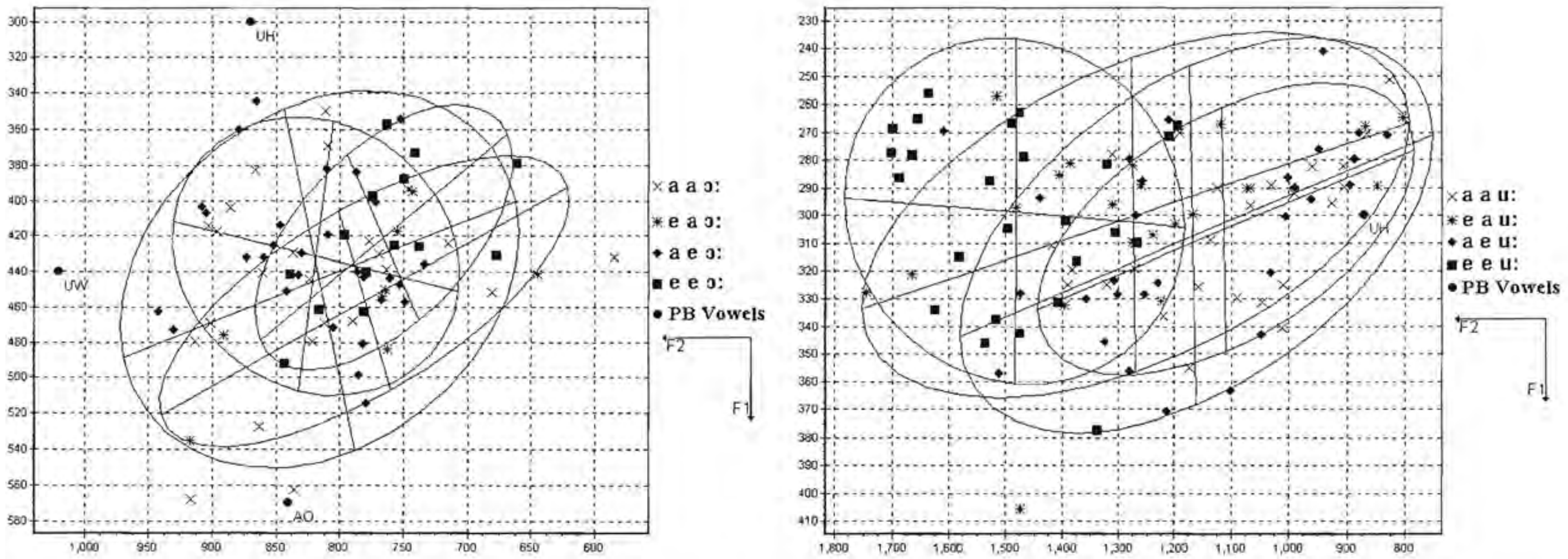


Figure A.6: Vowel formant clusters: [<ɔ:> in "*dom*" and in "bought"] and [<u:> in "*boer*" and "soon"]. The first *a/e* indicates the mother-tongue of the speakers and the second *a/e* indicates from which language the vowel was indicted as coming from.

Appendix

# Bibliography


[1] L.R. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, 1993.

[2] J.D. Markel and Jr. A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, 1982.

[3] E.C. Botha and L.C.W. Pols, "Modelling the acoustic differences between L1 and L2 speech: The short vowels of Afrikaans and South African English," in *Proceedings of the 5th European Conference on Speech Communication and Technology*, Rhodes, Greece, 1997, vol. 2, pp. 1035–1038.

[4] D. Jones (edited by P. Roach and J. Hartman), *English Pronouncing Dictionary*, Cambridge University Press, Cambridge, 1997.

[5] D.R. Miller and J. Trischitta, "Statistical dialect classification based on mean phonetic features," in *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, USA, 1996, vol. CDROM, p. none given.

[6] C. Teixeira, I. Trancoso and A. Serralheiro, "Accent identification," in *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, USA, 1996, vol. CDROM, p. none given.

[7] A.W.F. Huggins and Y. Patel, "The use of shibboleth words for automatically classifying speakers by dialect," in *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, USA, 1996, vol. CDROM, p. none given.

[8] V.V. Digalakis and G. Neumeyer, "Speakers adaptation using combined transformation and Bayesian methods," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 4, pp. 294–300, July 1996.

[9] D. Jones, *An Outline of English Phonetics*, W. Heffer and Sons, Cambridge, 1964.

[10] G. Peterson and H. Barney, "Control methods used in a study of vowels," *Journal of the Acoustic Society of America*, vol. 42, pp. 175–184, 1952.

[11] A. Holbrook and G. Fairbanks, "Diphthong formants and their movements," *Journal of Speech and Hearing Research*, vol. 5, no. 1, pp. 38–58, 1962.

[12] J.R. Taylor and J.Z. Uys, "Notes on the Afrikaans vowel system," *Leuvense Bijdragen*, vol. 77, no. 2, pp. 129–149, 1988.

[13] A. van der Merwe, E. Groenewald, D. van Aardt and H.E.C. Tesner, "Die formantpatrone van Afrikaanse vokale soos geproduseer deur manlike sprekers," *Suid Afrikaanse Tydskrif vir Taalkunde*, vol. 11, no. 2, pp. 71–79, 1993.

[14] H. Raubenheimer, "Enkele aspekte van die temporele eienskappe van lang vokale en diftonge in Afrikaans," M.S. thesis, Potchefstroom University for Christian Higher Education, 1994.

[15] H. Raubenheimer, *Acoustical features of diphthongs in Afrikaans*, Ph.D. thesis, Potchefstroom University for Christian Higher Education, 1998.

[16] M. Padmanabhan, L.R. Bahl, D. Nahamoo and M.A. Picheny, "Speaker clustering and transformation for speaker adaptation in speech recognition systems," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 1, pp. 71–77, 1998.

[17] C. Grover, D.G. Jamieson and M.B. Dobrovolsky, "Intonation in English, French and German: Perception and production," *Language and Speech*, vol. 30, no. 5, pp. 277–296, 1987.

[18] J.E. Flege, "The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification," *Journal of Phonetics*, vol. 15, pp. 47–65, 1987.

[19] R.A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Longman Group Ltd., London, 1964.

[20] I.C. Ward, *The Phonetics of English*, Heffer, Cambridge, 1958.

[21] E.C. Botha, "Towards modelling acoustic differences between L1 and L2 speech: The short vowels of Afrikaans and South-African English," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Amsterdam, The Netherlands, November 1996, vol. 20, pp. 65–80.

[22] H.F.V. Boshoff and E.C. Botha, "A new acoustic reference frame for vowels," in *Proceedings of the International Conference of Phonetic Sciences*, Berkeley, CA, USA, 1999, vol. Obtained from authors, p. none available.

[23] A.E. Coetzee, *Fonetiek vir eerstejaars*, Academica, Johannesburg, 1982.

[24] L.F. Willems, "Robust formant analysis," Tech. Rep. 529, Institute for Perception Research, Eindhoven, The Netherlands, April 1986.

[25] S.S. McCandless, "An algorithm for automatic formant extraction using Linear Prediction Spectra," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 22, no. 2, pp. 135–141, April 1974.

[26] P. Zolfaghari and T. Robinson, "Formant analysis using mixtures of Gaussians," in *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA, USA, 1996, vol. CDROM, p. none given.

[27] C. Snell and F. Milinazzo, "Formant location from LPC analysis data," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 2, pp. 129–134, April 1993.

[28] L. Welling and H. Ney, "Formant estimation for speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 1, pp. 36–48, January 1998.

[29] D. Delsarte and Y.V. Genin, "The Split Levinson Algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 3, pp. 470–478, June 1986.

[30] N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction," *Journal of Mathematics and Physics*, vol. 25, pp. 261–278, 1946.

[31] J.N. Holmes, W.J.Holmes and P.N. Garner, "Using formant frequencies in speech recognition," in *Proceedings of the 5th European Conference on Speech Communication and Technology*, Rhodes, Greece, 1997, vol. 3, pp. 2083–2086.

[32] J.H.L. Hansen and L.M. Arslan, "Foreign accent classification using source generated prosodic features," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, USA, 1995, IEEE, vol. 1, pp. 836–839.

[33] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Amsterdam, The Netherlands, 1993, vol. 17, pp. 97–110.

[34] J. Clark and C. Yallop, *An Introduction to Phonetics and Phonology*, Blackwell, Oxford, 1990.

[35] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg and C.A. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 399–418, October 1976.

[36] J.E. Flege, M.J. Munro and I.R.A. MacKay, "Factors affecting strength of perceived foreign accent in a second language," *Journal of the Acoustic Society of America*, vol. 5, no. 1, pp. 3125–3134, May 1995.

[37] J.H. Mathews, *Numerical Methods for Mathematics, Science, and Engineering*, Prentice Hall, New Jersey, 1992.

[38] M.R. Spiegel, *Probability and Statistics*, McGraw-Hill, Singapore, 1980.

[39] H.J. Rousseau, *Die Invloed van Engels op Afrikaans*, Miller, Cape Town, 1937.

[40] M. de Villiers and F.A. Ponelis, *Afrikaanse Klankleer*, Tafelberg, Cape Town, 1987.

[41] D.P. Wissing, *Algemene Afrikaanse en Generatiewe Fonologie*, Macmillan, Johannesburg, 1982.

[42] J.G.H. Combrink and L.G. de Stadler, *Afrikaanse Fonologie*, Macmillan, Johannesburg, 1987.

[43] D.R. van Bergem, "Perceptual and acoustical aspects of lexical vowel reduction, a sound change in progress," *Speech Communication*, , no. 16, pp. 329–358, 1995.

[44] P. van Alphen and D.R. van Bergem, "Markov models and their application in speech recognition," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Amsterdam, The Netherlands, 1989, vol. 13, pp. 1–26.