# Automatic Detailed Transcription of Speech Using Forced Alignment and Naïve Pronunciation Rules

**Per-Anders Jande**
**jande@speech.kth.se**
**CTT, TMH, KTH, GSLT**

**Abstract**

Automatically generated detailed transcriptions should be helpful for transcribers annotating speech databases with detailed transcriptions. In this paper, a method for creating detailed transcriptions using a digitised speech signal, word boundary annotation, canonical transcriptions and naïve pronunciation rules is described. Some initial results using word boundaries from an automatic alignment system are presented. Although this data has been based on partly erroneous alignments, the results seem plausible. The system will be developed so that it can use manually annotated word boundaries for forced alignment of canonical transcriptions. This will probably enhance system performance considerably.

## 1 Introduction

A general model for the pronunciation of words in context can be useful e.g. for increasing the naturalness of speech synthesis. Such a model could also be useful for enhancing speech recognition systems.

The goal of my research is to crate a human-readable rule system that takes a canonical phonological transcription (a transcription corresponding to a maximally detailed or near-maximally detailed pronunciation of a word, i.e. the kind of transcription normally found in a pronunciation lexicon) and adapts this transcription to the particular context described by a set of context parameters. Some candidate context parameters that may influence the pronunciation of a word are *part of speech*, *phrase types*, *phrase boundaries*, *prosodic boundaries* (and other linguistically important boundaries), *word and collocation frequencies*, *word length*, *information structure*, *discourse centrality*, *intonation*, *word stress and accent*, *disfluency context*, *hesitation context*, *local speech rate* (at word and/or syllable level), *speaking style* and *speaker attributes.*

Information about these parameters or information about parameter estimates will be used by machine learning algorithms to find suitable rules for the pronunciation of a word given the canonical phonetic transcription. The machine learning algorithms will also need the correct context-dependent – detailed – transcriptions (transcriptions corresponding to an actual uttered instance of a word) for performance evaluation. Manual transcription is laborious and if this process could be partly automated, much time and effort could be saved. Automatic support for detailed phonetic transcription is useful irrespective of the intended use of the transcriptions.

This paper will describe a system for automatically creating detailed transcriptions using the digitised speech signal, word boundary annotation, canonical transcriptions and naïve pronunciation rules.

### 1.1 Aim

The aim for the work discussed in this paper was to develop a system that can adapt a canonical transcription to a detailed transcription of an actual utterance by looking at the speech signal. The system cannot be expected to perform as well as phonetically trained human transcribers[1]. Manual correction of the system output will probably be necessary to get detailed transcriptions with the level of accuracy needed for the intended use of the transcriptions. However, the system transcriptions should have a level of accuracy which significantly lowers manual correction time compared to correcting canonical transcriptions directly (and correction should of course be faster than manual transcription without automatic support).

The system takes manually determined word boundaries and canonical transcriptions supplied by a forced recognition system as input. A *forced recognition* or *forced alignment* system knows the uttered word sequence in advance and the task of the system is to find the best alignment of the transcription sequence to the signal.

For each word, the system creates a state transition network describing the possible pronunciations of the word given a set of possible detailed realisations of each canonical segment (naïve pronunciation rules). Each state in the network will output a detailed segment or NULL (i.e., nothing).

The best path through the state transition network is calculated using a set of hidden Markov Models describing phonemes/segments (phoneme HMMs) and the part of the speech signal within the given word boundaries. The best path trough the network gives the optimal detailed transcription.

## 2 Method

### 2.1 Material

There are several speech databases available. However, the data discussed in this paper is based only on the the VaKoS database (Bannert and Czigler, 1999). This database consists of approximately 100 minutes of spontaneous speech in digital format from ten speakers of central standard Swedish. There is about ten minutes of monological speech on some predetermined topic from each speaker. The speech is segmented by hand on the word level and partly segmented on the phone level. There are also various other types of annotation.

### 2.2 Tools

A system for automatic time-aligned phonetic transcription developed by Kåre Sjölander (Sjölander, 2003) was used for generating time-aligned canonical transcriptions for the words in the speech data.

## 2.3 Phonetic Alphabet

The phonetic alphabet used by Sjölander's alignment system is close to the Swedish Transcription Alphabet (STA), originally developed for the RULSYS text-to-speech system (Carlson and Granström, 1976; Carlson et al., 1982). The set of phonemes is: {`P, p, T, t, K, k, B, b, D, d, G, g, F, V, S, SJ, TJ, H, M, N, NG, L, J, R, RT, RD, RL, RN, RS, EO, A, A:, E, E:, I, I:, O, O:, U, U:, Y, Y:, Å, Å:, Ä, Ä:, Ä4, Ä3, Ö, Ö:, Ö4, Ö3`}. The acoustic model used by the aligner includes separate models for the plosive occlusion phases ({`P, T, K, B, D, G, RT, RD`}) and the release/aspiration phases ({`p, t, k, b, d, g`}). In the canonical transcriptions, a plosive occlusion is always followed by its respective release. In the detailed output transcriptions, the release of a plosive is optional. A release/aspiration phase symbol can only follow its respective occlusion phase. It cannot stand alone or follow the occlusion of any other plosive. This is a phonological constraint which applies both to the canonical and to the detailed phonetic transcriptions. The consonant set {`RT, RD, RL, RN, RS`} denote reflexive consonants. The plosives `T` and `RT` share the release/aspiration model `t` and the plosives `D` and `RD` share the release/aspiration model `d`. The remaining phoneme symbols used correspond to symbols in STA (cf. Appendix A for a description of STA).

The detailed transcriptions can contain a `null` symbol, which is used as a place holder used for symbolic alignment between the canonical and the detailed transcriptions.

## 2.4 Canonical-to-Detailed Conversion

The detailed phonetic transcriptions are generated using an adaptation of a grapheme-to-phoneme mapper used by Sjölander's aligner (Sjölander, 2003). This mapper uses the acoustic input and a set of naïve grapheme-to-phoneme rules for mapping. The grapheme-to-phoneme mapper has been changed to perform canonical-to-detailed conversion using predetermined word boundaries. The canonical-to-detailed mapper thus uses a canonical phoneme string (aligned to the speech signal), a set of word boundaries, the speech sound file and Sjölander's phoneme HMMs, parameterisation procedure and speech alignment algorithm to estimate the best detailed transcription of the speech contained in the sound file.

The system also uses a table of possible output segments for each input segment (naïve canonical-to-detailed rules), including the possible deletion of each input segment. The system outputs two strings: 1) a detailed transcription with optimal alignment to the signal given the word boundaries and 2) a detailed transcription with the same alignment as the canonical transcription, using `null` symbols to denote deleted segments. The conversion table used by the program is shown in Appendix B. The entire plosive (occlusion phase and release/aspiration phase) is treated as one unit. Plosives without release/aspiration phases are treated as separate allophones. The plosive occlusion and release/aspiration always occur in pairs in the canonical string.

### 2.4.1 Generating a State Transition Network

From each word transcription, a set of indexed states is generated by the canonical-to-detailed conversion system. Each state can output a phonetic symbol corresponding to a phoneme HMM in the acoustic model or NULL. The state also outputs an alignment index which is used to map the detailed output transcription to the canonical transcription. The states are connected to a pronunciation network through the generation of a set of allowed transitions.

The possible transitions through the pronunciation network are determined in such a way that there is an arc from each non-NULL state with alignment index *n* to the NULL state with alignment index *n*. From this NULL state, there is an arc into each *n+1* state, except for states corresponding to a plosive release/aspiration phase; there is only an arc into the respective occlusion phase state. From the occlusion phase state, there is an arc into the release/aspiration phase state, forcing the model to go through each of the two states, if any. Since the release/aspiration phase is optional, there is a second occlusion phase added without a respective release/aspiration phase.

There could be an arc from the occlusion phase state into the release/aspiration phase state and arcs from both the occlusion phase state and the release/aspiration phase state into the NULL state with the same alignment index. This would give the same effect with one less state. However, having the representation with an extra state allows for forcing a certain plosive to always have a release by simply removing the second occlusion symbol from the conversion table; no change of the actual code is necessary. This representation is thus chosen for its flexibility.

| State | Symbol | Alignment index | Note |
|---|---|---|---|
| 0 | NULL | -1 | Initial state |
| 1 | J | 0 | |
| 2 | NULL | 0 | |
| 3 | O: | 1 | |
| 4 | O | 1 | |
| 5 | EO | 1 | |
| 6 | NULL | 1 | |
| 7 | RD | 2 | |
| 8 | d | 2 | |
| 9 | RD | 2 | |
| 10 | NULL | 2 | |
| 11 | EO | 3 | |
| 12 | NULL | 3 | Final state |

**Table 1.** States generated for the word *gjorde* (`JO:RDdEO`).

Table 1 shows the state set for the word *gjorde* (Eng. *did*) with the canonical transcription `JO:RDdEO`. State 7 and 8 in table 1 correspond to a plosive occlusion phase and a release/aspiration phase, respectively. State 9 corresponds to an occlusion only. In Table 2, it can be seen that there is an allowed transition from state 7 into state 8 and an allowed transition from state 8 into state 10. There is also an allowed transition from state 9 into state 10. There are allowed transitions from state 6 into state 7 and state 9, but not into state 8.

| From state | To state |
|---|---|
| 0 | 1 |
| 1 | 2 |
| 0 | 2 |
| 2 | 3 |
| 3 | 6 |
| 2 | 4 |
| 4 | 6 |
| 2 | 5 |
| 5 | 6 |
| 2 | 6 |
| 6 | 7 |
| 7 | 8 |
| 8 | 10 |
| 6 | 9 |
| 9 | 10 |
| 6 | 10 |
| 10 | 11 |
| 11 | 12 |
| 10 | 12 |

**Table 2.** Allowed state network transitions for the states in table 1.

There is always an arc from the NULL state with alignment index $n$ into the NULL state with alignment index $n+1$. Following such an arc results in a `null` symbol in the symbolic alignment output string (as mentioned, the system also outputs the detailed transcription aligned to the speech signal).

In short, the possible transitions go from an initial NULL state through one (or possibly two for plosives) or none of the possible realisations of the phoneme at position $n$ in the canonical transcription. The alignment index of a state refers to this position. The segment sequence corresponding to the path with the lowest score given the signal and the phoneme HMMs is returned. The score can be seen as the "distance" between the transcription and the signal.

Figure 1 shows a state transition graph illustrating the possible pronunciations of the word *gjorde* (`JO:RDdE0`) determined by the states shown in Table 1 and the transitions shown in Table 2.

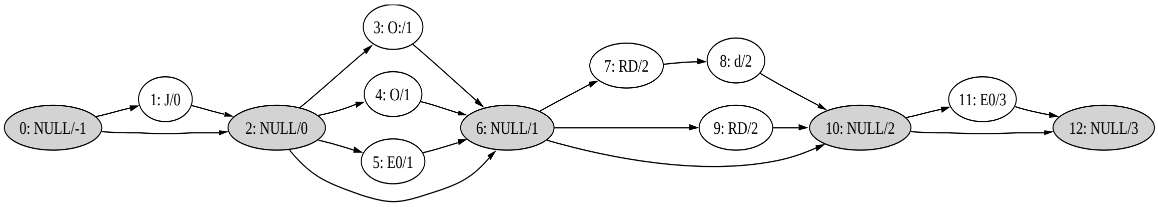

**Figure 1.** State transition graph for the word *gjorde* (`JO:RDdE0`).

### 2.5 Word Level Alignment

Using the word boundary labels included in the VaKoS database (Bannert and Czigler, 1999) as bounds for canonical segment alignment proved difficult, since the alignment is dependent on the acoustic input. It is not always possible for the aligner to fit an entire canonical transcription into a certain time frame, if many of the segments in the canonical string are not present in the signal. A version of the alignment system that can handle exceptional cases is under development. This aligner version uses the signal and the phoneme HMMs to do the alignment when possible, but when it is not possible to fit the canonical string to the signal using the phoneme HMMs, it will simply divide the space within the given word boundary into $N$ divisions of equal length, $N$ being the number of segments in the canonical string.

A division into equal length ranges for each phoneme when the alignment system cannot squeeze the phoneme string between the given word boundaries is not expected to cause any problems. The canonical phoneme boundaries are not used by the canonical-to-detailed converter. In fact, the canonical phoneme boundaries are mostly used for alignment purposes. For this, the exact positions of these boundaries are not critical. The canonical segment boundaries are used also for estimating some input parameters to machine learning systems, but the type of adjustments made will not be critical in that context either.

The alignment of the canonical transcriptions used so far is the output from Sjölander's aligner (Sjölander, 2003) without forcing it to fit the transcriptions within the manually determined word boundaries. The canonical-to-detailed system is then forced to use the alignment from the canonical transcription. This is a "backwards" way of doing things. However, it is only a very temporary solution to be able to test the method. The results of the canonical-to-detailed conversion are presently not very reliable, since the aligner has been forced to use transcriptions that do not fit the speech data very well and the canonical-to-detailed converter thus is forced to sometimes look at the wrong part of the speech signal. However, some initial results are presented in the following section.

## 3 Results

So far, there is no gold standard to match the result against for system performance evaluation. However, a gold standard will be compiled by letting human transcribers correct the canonical-to-detailed system output and the system will be evaluated against this gold standard[2].

Appendix C shows the possible input symbols to the canonical-to-detailed converter and the respective output symbols with frequency values for the initial test using automatically determined word boundaries. Although the word boundaries are not always correct, the initial results still look plausible. For example, short and/or central vowels are more prone to be transcribed as `E0` than long and/or peripheral vowels.

There were altogether 55,728 canonical phoneme symbols in the canonical transcriptions of the VaKoS database (Bannert and Czigler, 1999). Of these, 10,668 were converted to `null` symbols. There was thus an 19.1% reduction of the phonetic string.

### 3.1 Some Output Examples

Some more detailed examples of output from the canonical-to-detailed conversion system are shown in Table 4 and Table 5.

| Canonical | J | O: | RD | d | E0 |
|---|---|---|---|---|---|
| Detailed #1 | null | E0 | RD | | null |
| Detailed #2 | J | null | null | | E0 |
| Detailed #3 | J | O: | RD | | E0 |
| Detailed #4 | J | O | RD | | E0 |
| Detailed #5 | J | O | RD | d | null |
| Detailed #6 | J | E0 | RD | | E0 |
| Detailed #7 | J | E0 | RD | d | null |
| Detailed #8 | null | O | RD | d | E0 |
| Detailed #9 | J | null | RD | d | E0 |
| Detailed #10 | J | O: | RD | d | E0 |
| Detailed #11 | J | O: | RD | | null |
| Detailed #12 | J | E0 | RD | d | E0 |
| Detailed #13 | null | O: | RD | d | E0 |
| Detailed #14 | null | E0 | RD | d | E0 |

**Table 4.** The detailed transcriptions generated for the word *gjorde* (JO:RDdE0) in the VaKoS database (Bannert and Czigler, 1999).

Table 4 shows the altogether 14 different detailed transcriptions generated for the instances of the word *gjorde* (JO:RDdE0) in the VaKoS database (Bannert and Czigler, 1999) when automatically determined word boundaries were used. Of course, it is not possible to assess the accuracy of the transcriptions without listening to the corresponding speech. However, the detailed transcription with index 9 in Table 4 seems like an implausible pronunciation.

```
...
2.5600000 2.6100000 RD
2.6100000 2.7100000 E0
...
```

**Table 5.** Excerpt from an alignment file showing an instance of the detailed transcription with index 1 in table 4 with optimal alignment to the signal.

Table 5 shows an instance of the detailed transcription with index 1 in table 4 aligned to the speech signal using the optimal alignment of the detailed transcription to the signal (as determined by the conversion system). The numbers denote the segment boundaries (in seconds from the start of the sound file).

Table 6 shows the same transcription instance aligned to the signal using the segment boundaries determined by the aligner for the canonical transcription (note that the occlusion phase and the release/aspiration phase of the plosive are aligned to the speech signal as separate units.) The word boundaries are the same for both alignments – those given by the automatic alignment.

```
...
2.5600000 2.5900000 null
2.5900000 2.6200000 E0
2.6200000 2.6500000 RD
2.6500000 2.6800000 null
2.6800000 2.7100000 null
...
```

**Table 6.** Excerpt from an alignment file showing an instance of the detailed transcription with index 1 in table 4 with the alignment for the canonical transcription.

## 4 Conclusions

A system for creating detailed transcriptions using a digitised speech signal, word boundary annotation, canonical transcriptions and naïve pronunciation rules has been described. Some initial results using word boundaries from an automatic alignment system have been presented.

When listening to the speech signal and simultaneously looking at a spectrogram representation of the signal with the aligned word boundaries and the canonical-to-detailed system output, it is obvious that the system does well when the aligner has done well, but that the converter does poorly when the aligner has done poorly (and the converter thus is looking at the wrong part of the signal). Although presently unknown, the system performance is expected to increase considerably when the manually segmented word boundaries are used. However, after an informal inspection of the output, it is judged that this kind of automatic detailed transcription can be a good help for transcribers when annotating speech data with detailed phonetic transcriptions.

## 5 Future Work

As mentioned, the aligner will have to be adapted to handle cases where the word is too short for the phonetic string. This work is in progress. It has also been mentioned that a gold standard will have to be compiled, so that the system can be properly evaluated. Since the transcriptions are to be used in machine learning, a high degree of accuracy is necessary. It is not expected that the system will give an output that can be used directly, but the output is thought to be a support for manual transcription. Thus, all canonical-to-detailed output will probably have to be corrected by hand.

The orthographic transcriptions of all available databases will have to be aligned to the speech signal before the system can be used for these databases – the VaKoS database (Bannert and Czigler, 1999) is the only database for which this has been done. The alignment system developed by Sjölander (Sjölander, 2003) can be used to facilitate this work, but the output will have to be manually checked and corrected.

Finally, it would be interesting to train a set of more detailed phoneme models, including a larger set of allophones (perhaps including some devoiced plosives and some nasalised vowels). This would require re-labelling speech databases using the enlarged phoneme inventory to create training data.

## Acknowledgements

Many thanks to Kåre Sjölander for acoustic models, code and help. Many thanks also to Robert Bannert and Peter Czigler for making their VaKoS database available. The research reported in this paper was conducted at the Centre for Speech Technology (CTT) at KTH.

## Notes

[1] There is no objective correct answer for this kind of task and the inter-transcriber agreement is expected to be far from perfect. Although humans make occasional errors, the human transcribers are expected to supply transcriptions that are all reasonable in the sense that they cannot be considered to be incorrect, only different. Even if the result of the automatic detailed transcription system does not stand out when transcriptions are evaluated for inter-transcriber agreement, it cannot be assumed that the system transcriptions are reasonable in the same way as the human transcriptions.

[2] This will probably result in a higher conversion accuracy than using a completely manually transcribed gold standard. This should be kept in mind if comparisons are made to evaluations made on completely manual gold standard transcriptions.

## References

Bannert, R. and P. E. Czigler (1999). *Variations in consonant clusters in standard Swedish.* Phonum 7, Reports in Phonetics. Umeå: Umeå University.

Carlson, R. and B. Granström (1976). A text-to-speech system based entirely on rules. In *Proceedings of ICASSP*, pp. 686–688.

Carlson, R., B. Granström, and S. Hunnicutt (1982). A multi-language text-to-speech module. In *Proceedings of ICASSP*, volume 3, pp. 1604–1607.

Sjölander, K. (2003). An HMM-based system for automatic segmentation and alignment of speech. In *Procceedings of Fonetik 2003*, pp. 93–96.

# Appendix A – The Swedish Transcription Alphabet (STA)

| STA | Explanation | Example word | Example transcription |
|---|---|---|---|
| ' | Accent I stress marker | gå | G'Å: |
| " | Accent II main stress | gata | G"A:TA |
| ‘ | Accent II secondary stress in compounds | gågata | G"Å:hyG‘A:TA |
| hy | Compound boundary | gågata | G"Å:hyG‘A:TA |

**Table A1.** STA accent markers and compound boundary marker with explanations and examples in Swedish.

| STA | IPA | Example word | Example transcription |
|---|---|---|---|
| A: | ɒː | lam | L'A:M |
| A | a | lamm | L'AM |
| E: | eː | se | S'E: |
| E | e | sett | S'ET |
| E0 | ə | natten | N'ATEON |
| I: | iː | sil | S'I:L |
| I | ɪ | sill | S'IL |
| O: | uː | bo | B'O: |
| O | ʊ | bott | B'OT |
| U: | ʉː | hus | H'U:S |
| U | ɵ | hund | H'UND |
| Y: | yː | myra | M"Y:RA |
| Y | ʏ | mygga | M"YGA |
| Å: | oː | gå | G'Å: |
| Å | ɔ | gått | G'ÅT |
| Ä: | æː | häl | H'Ä:L |
| Ä | ɛ | sätt | S'ÄT |
| Ä3 | æ̞ː | här | H'Ä3R |
| Ä4 | ɛ̞ | herr | H'Ä4R |
| Ö: | øː | öga | "Ö:GA |
| Ö | ø | löss | L'ÖS |
| Ö3 | œː | öra | "Ö3RA |
| Ö4 | œ | förr | F'Ö4R |

**Table A2.** STA vowel symbols with IPA equivalents and examples in Swedish.

| STA | IPA | Example word | Example transcription |
|---|---|---|---|
| B | b | bar | B'A:R |
| D | d | bod | B'O:D |
| 2D | ɖ | bord | B'O:2D |
| F | f | far | F'A:R |
| G | g | gul | G'U:L |
| H | h | hatt | H'AT |
| J | j | jul | J'U:L |
| K | k | tack | T'AK |
| L | l | lus | L'U:S |
| 2L | ɭ | sorl | S'Å:2L |
| M | m | mor | M'O:R |
| N | n | natt | N'AT |
| 2N | ɳ | barn | B'A:2N |
| NG | ŋ | tung | T'UNG |
| P | p | pil | P'I:L |
| R | r | ros | R'O:S |
| S | s | sol | S'O:L |
| SJ | ɕ, ɧ | skjorta | SJ"ORTA |
| 2S | ʂ | dusch | D'U2S |
| T | t | till | T'IL |
| 2T | ʈ | hårt | H'Å:2T |
| TJ | ʃ | kedja | TJ"E:DJA |
| V | v | var | V'A:R |

**Table A3.** STA consonant symbols with IPA equivalents and examples in Swedish.

# Appendix B – Canonical-to-Detailed Conversion Table

| Input segment | | Possible output segments |
|---|---|---|
| P p | ⟶ | null, P p, P, B b, B |
| T t | ⟶ | null, T t, T, D d, D, RT t, RT |
| K k | ⟶ | null, K k, K, G g, G |
| B b | ⟶ | null, B b, B, P p, P |
| D d | ⟶ | null, D d, D, T t, T, RD d, RD |
| G g | ⟶ | null, G g, G, K k, K |
| F | ⟶ | null, F, V |
| V | ⟶ | null, V, F |
| S | ⟶ | null, S, RS |
| SJ | ⟶ | null, SJ |
| TJ | ⟶ | null, TJ |
| H | ⟶ | null, H |
| M | ⟶ | null, M |
| N | ⟶ | null, N, M, NG, RN |
| NG | ⟶ | null, NG |
| L | ⟶ | null, L, RL |
| J | ⟶ | null, J |
| R | ⟶ | null, R |
| RT t | ⟶ | null, RT t, RT |
| RD d | ⟶ | null, RD d, RD |
| RL | ⟶ | null, RL |
| RN | ⟶ | null, RN |
| RS | ⟶ | null, RS |
| EO | ⟶ | null, EO |
| A | ⟶ | null, A, EO |
| A: | ⟶ | null, A:, A, EO |
| E | ⟶ | null, E, EO |
| E: | ⟶ | null, E:, E, EO |
| I | ⟶ | null, I, EO |
| I: | ⟶ | null, I:, I, EO |
| O | ⟶ | null, O, EO |
| O: | ⟶ | null, O:, O, EO |
| U | ⟶ | null, U, EO |
| U: | ⟶ | null, U:, U, EO |
| Y | ⟶ | null, Y, EO |
| Y: | ⟶ | null, Y:, Y, EO |
| Å | ⟶ | null, Å, EO |
| Å: | ⟶ | null, Å:, Å, EO |

...

| Input segment | | Possible output segments |
|---|---|---|
| ... | | |
| Ä | ⟶ | null, Ä, E0 |
| Ä: | ⟶ | null, Ä:, Ä, E0 |
| Ä4 | ⟶ | null, Ä4, E0 |
| Ä3 | ⟶ | null, Ä3, Ä4, E0 |
| Ö | ⟶ | null, Ö, E0 |
| Ö: | ⟶ | null, Ö:, Ö, E0 |
| Ö4 | ⟶ | null, Ö4, E0 |
| Ö3 | ⟶ | null, Ö3, Ö4, E0 |

**Table B1.** Canonical-to-detailed conversion table.

# Appendix C – Canonical-to-Detailed Conversion Results

| Input | | Output symbol/frequency |
|---|---|---|
| P p | ⟶ | P p/734, null/30, P/28, B b/9, B/3 |
| T t | ⟶ | T t/3461, RT t/190, null/155, T/96, D d/78, RT/22, D/21 |
| K k | ⟶ | K k/1898, null/91, K/91, G g/50, G/19 |
| B b | ⟶ | B b/336, P p/142, null/84, B/84, P/19 |
| D d | ⟶ | T t/758, D d/670, null/325, D/131, T/102, RD/90, RD d/85 |
| G g | ⟶ | G g/271, K k/252, G/107, null/94, K/43 |
| F | ⟶ | F/997, null/41, V/32 |
| V | ⟶ | V/882, null/499, F/250 |
| S | ⟶ | S/2165, RS/761, null/355 |
| SJ | ⟶ | SJ/172, null/15 |
| TJ | ⟶ | TJ/76, null/9 |
| H | ⟶ | H/871, null/262 |
| M | ⟶ | M/1691, null/538 |
| N | ⟶ | N/1856, null/970, NG/437, M/360, RN/327 |
| NG | ⟶ | NG/326, null/109 |
| L | ⟶ | L/1578, null/641, RL/77 |
| J | ⟶ | J/858, null/468 |
| R | ⟶ | R/2167, null/1622 |
| RT t | ⟶ | RT t/112, RT/36, null/15 |
| RD d | ⟶ | RD d/43, RD/41, null/13 |
| RL | ⟶ | null/1 |
| RN | ⟶ | RN/62, null/33 |
| RS | ⟶ | RS/138, null/13 |
| E0 | ⟶ | E0/1814, null/678 |
| A | ⟶ | E0/2153, A/1656, null/783 |
| A: | ⟶ | A:/581, E0/388, A/275, null/181 |
| E | ⟶ | E0/1115, null/405, E/376 |
| E: | ⟶ | E0/673, E:/510, null/246, E/230 |
| I | ⟶ | I/1099, E0/476, null/447 |
| I: | ⟶ | I:/404, I/261, E0/169, null/92 |
| O | ⟶ | O/134, E0/76, null/75 |
| O: | ⟶ | O:/186, O/124, E0/46, null/38 |
| U | ⟶ | E0/259, null/216, U/188 |
| U: | ⟶ | U:/275, E0/82, null/65, U/9 |
| Y | ⟶ | Y/139, null/76, E0/72 |
| Y: | ⟶ | Y:/28, E0/11, Y/6, null/3 |
| Å | ⟶ | Å/965, E0/826, null/414 |
| Å: | ⟶ | E0/661, Å:/591, Å/280, null/202 |

...

| Input | | Output symbol/frequency |
|---|---|---|
| ... | | |
| Ä | ⟶ | E0/370, Ä/116, null/110 |
| Ä: | ⟶ | E0/131, Ä:/43, Ä/40, null/38 |
| Ä4 | ⟶ | E0/44, Ä4/33, null/14 |
| Ä3 | ⟶ | E0/302, Ä3/103, null/58, Ä4/26 |
| Ö | ⟶ | E0/49, Ö/21, null/10 |
| Ö: | ⟶ | E0/108, Ö:/73, Ö/29, null/15 |
| Ö4 | ⟶ | E0/67, null/61, Ö4/51 |
| Ö3 | ⟶ | E0/121, Ö3/68, null/58, Ö4/22 |

**Table C1.** Input symbols to the canonical-to-detailed conversion system and their sets of output symbols with frequencies.