Squibs and Discussions
Aligning Phonetic Segments for Children’s Articulation Assessment

Harold L. Somers*
UMIST

In a recent paper published in this journal (Covington 1996), an algorithm is described which aligns segments within a pair of words for the purpose of identifying historical cognates. This algorithm could have a further application in the field of speech therapy, and in particular in the practice of articulation assessment of young children. The present author developed a similar algorithm some years ago for this purpose. In this paper, we explore some points of comparison between the two applications.

1. Articulation Testing

It is well known that a child’s acquisition of phonology is gradual, and can be charted according to the appearance of phonetic distinctions (e.g., stops vs. fricatives), the disappearance of childish mispronunciations, especially due to assimilation ([9o9] for dog), and the ability to articulate particular phonetic configurations (e.g., consonant clusters). Childhood speech impediments, often a symptom of other learning disorders, can often be diagnosed early on by the identification of delayed acquisition of these articulatory skills. Whether screening whole populations of children, or assessing individual referrals, the articulation test is an important tool for the speech clinician.

A child’s articulatory development is usually described with reference to an adult model, and in terms of deviations from it: a number of phonological “processes” (Table 1) can be identified (see Ingram [1976]), and their significance with respect to the chronological age of the child assessed (though often processes interact, so for example when spoon is pronounced [fun] we have consonant-cluster reduction and assimilation). In Somers (1978a, 1979) I reported a computer program called CAT (Computerised Articulation Test), which I had developed to automate the assessment of children’s articulation tests. At the heart of this program was an algorithm very similar to the one reported by Covington.

Whereas Covington seeks to align the segments in possible historical cognates, CAT aligns the segments of a child’s articulation with those of the adult model, and on the basis of this looks for evidence of the phonological processes listed in Table 1. For example, if elephant [elɪfənt] is pronounced [ɛvət], we need to decide which of several possible alignments is the most plausible (cf. Covington 1996, 481):

\[
\begin{align*}
\text{ɛ-\text{-vət}} & \quad \text{ɛvət--} & \quad \text{ɛvə---t} & \quad \text{ɛ-\text{-vət-}} & \quad \text{ɛlɪfənt} & \quad \text{ɛlɪfənt} & \quad \text{ɛlɪfənt} & \quad \text{ɛlɪfənt}
\end{align*}
\]

If applied to a body of articulation data, e.g., a corpus of, say, 45 words elicited from the child as in a standardized articulation test, the evidence of each phonological process can be quantified, and the overall picture compared with the model of the average child, to give an individual’s “articulation age” and profile.

* Centre for Computational Linguistics, UMIST, Manchester, England. E-mail: harold@ccl.umist.ac.uk

© 1999 Association for Computational Linguistics
Table 1
Phonological processes.

| Process                        | Example | Adult Model | Child’s Version |
|--------------------------------|---------|-------------|-----------------|
| Final consonant deletion       | queen   | kwin        | kwi             |
| Unstressed syllable deletion   | elephant| elfont      | cfont           |
| Consonant cluster reduction    | stamps  | stamps      | tam             |
| Stopping                       | kiss    | kis         | kit             |
| Fronting                       | key     | ki          | ti              |
| Denasalisation                 | mummy   | mAmi        | bAbi            |
| Affrication                    | tent    | tent        | tent            |
| Vocalisation (1)               | bottle  | botl        | botu            |
| Vocalisation (2)               | chimney | tjImni      | tjImni          |
| Depalatalisation               | fish    | fis         |                 |
| Devoicing                      | dogs    | dogz        | doks            |
| Voicing                        | tent    | tent        | dent            |
| Assimilation                   | dog     | dog         | gog             |
| Lisping                        | kiss    | kis         | kîð             |
| S-lateralisation               | fish    | fis         | fil             |
| Ejectivisation                 | tent    | tent        | tent’           |
| Metathesis                     | remember| rimɛmb ø   | minɛmb ø        |
| Gliding                        | look    | luk         | wuk             |

2. Use of the Computer by Speech Clinicians

Early studies reporting the use of computers by speech pathologists include Faircloth (1971), van Demark and Tharp (1973), and Telage (1980), none of which involves automatic analysis of the input, though the last named uses binary articulatory features in a way almost identical to CAT. Comparatively little has appeared in the speech-language disorders literature on the specific topic of computerized articulation testing in the nearly 20 years since the CAT program was developed. Software for computerized language analysis does exist, but is mainly for grammatical and lexical analysis. Fairly thorough overviews are given by Rushakoff (1984), Rushakoff and Schwartz (1986), Long (1991), Long and Masterson (1993), and Miller and Klee (1995), though of course these may be more or less out of date. Other very general works such as Schwartz (1984), Curtis (1987), Silverman (1987), Cochran and Masterson (1995), and Masterson (1995) cover the use of computers by clinicians for all aspects of their work, including screening and diagnosis of various language skills (lexis, grammar, understanding, and auditory skills, as well as articulation) but also research (use of statistics), treatment (computer-based games), and clerical uses.¹

Those programs reported in the literature which specifically address the problem of articulation testing are listed below. For many of these programs, it seems that the only published information is in the user manual that accompanies the software. As far as one can tell, in none of the packages is the data analysis fully automatic. The following packages have been reviewed or discussed in the articles as cited: CAPP (Computer Analysis of Phonological Processes) (Long 1991; Kennedy 1986); Computer Managed Articulation Diagnosis (Bardzik 1986; Long 1991); Computerized Profiling 1.0 (Klee and Sahlie 1994); Computerized Profiling 2.0 (Gregg and Andrews 1995); ISPA

¹ Several papers refer to articles in the Journal for Computer Users in Speech and Hearing, but at the time of writing I was unfortunately unable to locate any copies of this obviously relevant journal.
The best-known computer application in speech-language pathology research is the CHILDES database of language samples and associated software (MacWhinney and Snow 1985; MacWhinney 1992; Sokolov and Snow 1994). This is primarily aimed at facilitating the storage and search of large databases of transcribed clinical data, where the transcription is basically orthographic, with mark-up for gestures, pauses, and other conversational features. Provision is made for a phonetic transcription too, using a “translation” of the IPA (International Phonetic Association) alphabet into ASCII symbols called “UNIBET” (MacWhinney 1992, 61ff). Although the organizers of the CHILDES database have had input from computational linguists on the question of mark-up, there is little or no automatic analysis. Crucially, no attempt is made to compare on a phone-by-phone basis the child language data with adult models, so data on the types of phonological process listed in Table 1 cannot be extracted.

This situation is typical of child language software, exemplified by Pye and Ingram (1988), whose PAL system uses a simple transcription, based on the IPA consonant chart, without the possibility of diacritics or special symbols to indicate specifically childish articulations. The system is unable to compare adult models with the child’s output, and can only produce a “phonological lexicon”, i.e., a list of the different sounds attested: it is then up to the clinician to analyze this inventory, e.g., to see if sounds are used contrastively, or in complementary distribution. The authors suggest that matching the child’s utterances to an adult model would involve a procedure which “would have to be very sophisticated indeed to handle complex cases of metathesis and deletion” (p. 124). As we show in Somers (1979) and in the next section, CAT was able to handle metathesis and deletion without being “very sophisticated indeed.”

3. The CAT Algorithm

Since the Somers (1979) article was aimed at speech therapists, it did not describe the alignment algorithm as such, which is described only in a local journal (Somers 1978a) and—in great detail—in an M.A. thesis (Somers 1978b). It bears comparison with Covington’s algorithm, though it should be said that the implementation in Pascal would be judged crude in the light of modern programming practice.

3.1 Coding the Input

The articulation data is coded as a narrow transcription, identifying phonetic detail such as secondary articulations, which can be important in speech therapy, in a fairly transparent notation, despite the limitations of the (capitals only) character set: primary phones are identified by single characters, with diacritics indicated in brackets, for example N(D) would indicate a dental (rather than alveolar) [n]. The notation is interpreted internally as bundles of articulatory features. The adult models are stored in a similar form. In each word, one vowel is marked as the primary stress, and this is taken as an anchor point for the alignment. The program as a whole ignores vowel quality, and in the CAT transcription any one of five vowel symbols (the vowel characters AEIOU in fact) can be used, the choice of one or the other being merely cosmetic. This treatment of vowels is a reasonable expedient. Primarily, CAT is aimed at consonant articulation, which is also the main concern of speech clinicians: see Stoel-Gammon and Herrington (1990) who state that “vowels are mastered earlier [than
consonants] and tend to evidence fewer errors" (p. 145). Regarding the identification of a single stressed vowel as an anchor point for the alignment, stress patterning (at least in stress-timed languages like English) is one of the first features of phonology to be acquired by children: again, if this is still a problem, then the distinctions tested by CAT will certainly be too fine-grained for such a subject. 2

3.2 Alignment
The alignment is based on taking the highest-scoring matches in terms of features, much as suggested by Covington (p. 490), so that in the elephant example above, the alignment [v]:[f] is preferred over the alignment [v]:[l]. Since the number of features for each segment remains constant, it is a simple matter of adding up the number of common features (+ or -), and taking the highest total. The algorithm works on the basis of “syllables” centred around a vowel. With the stressed vowel as an anchor point, the search-space is reduced to a comparison of the syllables either side of it: note that “vowel” is also marked as a feature. This is generally straightforward if the words are mono- or disyllabic, or trisyllabic with the stress on the second syllable. In other cases, if there is gross distortion of the consonants as well as inserted or omitted syllables, alignment can become somewhat arbitrary.

The algorithm takes some other factors into account, and is “tuned” to look out for certain processes that undermine the simplistic sequential skip-and-match search (which is also the basis of Covington’s algorithm): two such processes are metathesis (e.g., remember:[mIr~mb~]) and merging. In merging, a consonant cluster is simplified so that the resulting phone shares features of the two merged phones, e.g., box [boks] pronounced [bot], where the [t] has the place of articulation of the [s], but the manner of articulation of the [k]. Identifying metathesis can be rendered more complex by the coincidence of some other process, e.g., stopping, so that elephant becomes [cplIant] with the [l] and [f] swopped round, and the [f] replaced by a [p]. The CAT alignment algorithm looks for these explicitly. Figure 1 shows the result of the alignment of stamps:[daI3x], as it was actually presented.

The algorithm first aligns the marked vowel. It then takes the sequence of segments either side of the vowel. For [st]:[d], [d] is aligned with [t] rather than [s] as [t]:[d] represents a difference of only one feature, while [s]:[d] differs in three features. The evidence for a merge is the same as for the simpler devoicing analysis,

---

2 I am grateful to the anonymous reviewer who queried this aspect of the algorithm.
so the latter is preferred. In the case of \([mp]:[\sim]\), the algorithm compares the four possibilities \([m]:[\sim]\), \([p]:[\sim]\), \([p]:[x]\), and \([s]:[x]\), as well as the possible merges \([mp]:[\sim]\) and \([ps]:[x]\), in that order. It does not consider the matches \([m]:[x]\) or \([s]:[\sim]\) as these would involve a simultaneous insertion and deletion (cf. Covington's "no-alternating-skips rule," p. 482). As usual, the solution with the least "cost" in terms of feature differences is chosen. If the sequence includes a "no cost" match, this would immediately be preferred. The test for metathesis would also be made when there is an "unstressed" vowel in the sequence, though not in consonant clusters (so \(vest:[vets]\) would be analyzed incorrectly).

3.3 Comparing CAT and Covington's Algorithm

Comparing the CAT algorithm with Covington's, it seems that a key difference is the manual identification of a favored segment—the "stressed" vowel—as an anchor point. This can drastically reduce the search-space, especially if it happens to occur near the middle of the string, as in the above example. Apart from this, both algorithms work on a sequential match-or-skip, comparing the relative cost of each match, and narrowing the search-space by halting the search if a perfect match is found. The CAT algorithm has the additional task of searching explicitly under certain circumstances for metathesis and merges.

Apart from Covington's more sophisticated programming style, the only other difference between our techniques is in the scoring method. Covington's (p. 487) seems simpler than my own, in that the penalties reflect different types of (mis-)match, whereas in CAT the score derives more directly from the phonetic nature of the match. Covington states, on the same page, that "excessively narrow phonetic transcriptions do not help; they introduce too many subtle mismatches that should have been ignored." The CAT alignment algorithm, however, makes quite the opposite assumption, since the nature of the task demands a particularly narrow transcription. Covington also states (p. 490) that his algorithm could be improved by using phonetic features.

It is enlightening to take Covington's cognate alignment examples and to see what CAT would make of them. Looking first at the Spanish-French pairs (pp. 488f), we find that CAT agrees with Covington in 16 of the 20 cases. CAT has problems in three cases where the French has lost syllables that are stressed in Spanish, as in \(cabez:cap\) (1); \(drbo:harbre\) (2), CAT gets the correct alignment as identified by Covington (p. 488) if we omit the schwa in the French transcription (as would be normal for Parisian French (Armstrong 1967, 117).

Example 1

(1) \(k\ a\ b\ e\ \theta\ a\) \(k\ a\ b\ \bar{e}\ \theta\ a\)
    \(k\ a\ p\ -\ -\) \(k\ -\ -\ \bar{a}\ p\ -\)

(2) \(a\ r\ b\ -\ o\ 1\) \(\bar{a}\ r\ b\ -\ o\ 1\)
    \(a\ r\ b\ \bar{r}\ \bar{o}\) \(\bar{a}\ r\ b\ -\ -\ R\)

For the English-German data (pp. 490f) CAT gets exactly the same alignments as Covington for all 20 pairs (including the incorrect analysis of \(this:dises\)), though in CAT we would not transcribe the second element of the diphthongs in four of the examples. Like Covington's algorithm, CAT would correctly assign the \([\theta]\) of \(mouth\) with the \([t]\) rather than the \([n]\) of \(Mund\).

---

3 In this and subsequent examples, the CAT alignments (on the right) are shown in IPA; an acute accent marks the "stressed" vowel. Covington's alignments are shown on the left.
The examples considered so far have been quite straightforward (and much easier to align than typical child language data). The English-Latin cognates (pp. 492f) present more of a challenge. Applying the accepted rules of Latin stress, the CAT and Covington alignments differ in five of the 20 cases: In four of these, blow:flāre, fish:piscis, full:plēnus, and tooth:dentis (3), CAT does better than Covington, and in three other cases (grass:grāmen, heart:cordis, and mountain:mōns), CAT gets as first choice the alignment Covington ranks third, second, and second respectively. With just one exception (knee:genü), CAT does as well as or better than Covington.

Example 2

(3) --- tuwθ tū---θ
dentis dēntis

On the Fox-Menomini data (p. 494), CAT gets the same results as Covington on all ten examples if we assume either the first or the second vowel is stressed. Finally, Covington presents a variety of language-pair examples (p. 495). Again, the correct placement of the stressed vowel is important, leading to a wrong alignment for centum:hekaton (4), and preventing the [g]:[x] alignment in thugātēr:Tochter (5). CAT does worse than Covington in one other case didōmi:dō (6), but better in three cases daugh-ter:thugatēr (7), ager:ajras (8), and bharāmi:pherō (9). For centum:satem they both get the same alignment.

Example 3

(4) ---kentum --k---ēntum
hekaton hekatōn--
(5) thugatēr thugā-ter
tox-tor t--ōxter
(6) didōmi didōmi
--dō-- d--ō--
(7) ---dotēr d---ōtēr
thugatēr thugā-ter
(8) a-ger áger--
ajras ádjas--

In summary then, CAT does worse on the Spanish-French, better on the English-Latin, and about the same on the rest. Considering that Covington’s algorithm is aimed at dealing with this sort of data, this is a good result for CAT.

In a reciprocal comparison, Michael Covington was kind enough to run his algorithm on some child language data that I sent him. Of 25 examples, all of which CAT handles correctly, Covington’s algorithm also got the correct alignment, but often it was unable to distinguish between alternative alignments, all of which received the same score. For example, with the stamps:[da]jxs alignment mentioned above, all six different combinations of consonant alignment either side of the vowel are proposed with an equal score. This is because, as Covington (personal communication) readily points out, “it doesn’t know anything about place of articulation.”
4. Conclusions

4.1 Connolly's New Algorithm
Since the appearance of Covington's article (and even since the first draft of this reply), a highly relevant article has appeared, which—coincidentally—addresses the issues raised here (Connolly 1997). In this two-part article, Connolly first suggests ways of quantifying the difference between two individual phones, on the basis of perceptual and articulatory differences, and using either a Euclidean distance metric or, like CAT, a feature-based metric. Connolly's proposals are more elaborate, however, in that they permit specific differences to be weighted, so as to reflect the relative importance of each opposition. In the second part of the article, Connolly introduces a distance measure for comparing sequences of phones, based on the Levenshtein distance well-known in the speech processing and corpus alignment literature (inter alia). Again, this metric can be weighted, to allow substitutions to be valued differentially (presumably on the basis of the individual phone distance measure as described in the first part), and to deal with merging and metathesis. Connolly also considers briefly the effects of nonlinear prosodic structure on the distance measure. Although his methods are clearly computational in nature, Connolly reported (personal communication, 1997) that he had not yet implemented them. Taken together, these measures are certainly more sophisticated than either CAT's or Covington's, so this contribution could well be an extremely significant one towards the development of articulation testing software. In Somers (1998), I report an implementation and comparison of Connolly's measures with my own earlier work.

4.2 What Would a New Version of CAT Be Like?
In the light of the above remarks, it is interesting to think about how we might specify a reimplementation of CAT. One area where there could be considerable improvement is in the data input. CAT uses a very crude phonetic transcription based only on a minimal character set, not even including lower-case letters. Clearly this restriction would not be necessary nowadays. The software system PDAC (Phonological Deviation Analysis by Computer) uses a software package called LIPP (Logical International phonetic Programs) for input of transcriptions (Perry 1995). Alternatively, it seems quite feasible to allow the transcriptions to be input using a standard word processor and a phonetic font, and to interpret the symbols accordingly. For a commercial implementation it would be better to follow the standard proposed by the IPA (Esling and Gaylord 1993), which has been approved by the ISO, and included in the Unicode definitions.

Despite the reservations of all the speech-language pathology experts, it seems to me that the work on alignment discussed here (Somers 1978b; Covington 1996; Connolly 1997) suggests that this aspect of computerized articulation test analysis is a research aim well worth pursuing, especially if collaborators from the speech-language pathology field can be found. It would be rewarding if this article were to awaken interest in the problem.

Acknowledgments
I would like to thank the following people for their help in gathering the information presented in this paper: Catherine Adams (University of Manchester), Lawrence Shriberg (University of Wisconsin-Madison), Julie Masterson (Southwest Missouri State University), Carol Stoel-Gammon (University of Washington) and John Connolly (Loughborough University); Michael Covington, for collaborating on the "bake-off"; Joe Somers, for providing some of the example data; and the three anonymous reviewers for their suggestions, which have been extremely valuable.
References
Armstrong, Lilias E. 1967. The Phonetics of French: A Practical Handbook. G Bell, London.
Ball, Martin J. 1994. Review of ISP: Interactive System for Phonological Analysis. Child Language Teaching and Therapy, 10:97–102.
Bardzik, Paul P. 1986. Review of Computer Managed Articulation Analysis. ASHA, 28(2):74–75.
Cochran, Paula S. & Julie J. Masterson. 1995. NOT using a computer in language assessment/intervention: In defense of the reluctant clinician. Language, Speech, and Hearing Services in Schools, 26:213–222.
Connolly, John H. 1997. Quantifying target-realization differences. Clinical Linguistics & Phonetics, 11:267–298.
Covington, Michael A. 1996. An algorithm to align words for historical comparison. Computational Linguistics, 22:481–496.
Curtis, Jack F. 1987. An Introduction to Microcomputers in Speech, Language, and Hearing. College-Hill, Boston, MA.
Dyson, Alice T. 1987. Review of PEPPER: Programs to Examine Phonetic and Phonological Evaluation Records. Child Language Teaching and Therapy, 3:329–335.
Esling, John H. & Harry Gaylord. 1993. Computer codes for phonetic symbols. Journal of the International Phonetic Association, 23:83–97.
Faircloth, Marjorie A. 1971. Computer-assisted articulation analysis. Paper presented at the Annual Convention of the American Speech and Hearing Association, Chicago, IL, Sept. 1971. ASHA, 13:534.
Glegg, Ellen Meyer. & Valorie Andrews. 1995. Review of Computerized Profiling (1993). Child Language Teaching and Therapy, 11:209–216.
Ingram, David. 1976. Phonological Disability in Children. Edward Arnold, London.
Kennedy, Kathleen A. 1986. Review of Computer Analysis of Phonological Processes. ASHA, 28(8):71.
Klee, T. E. & Catherine Sahlie. 1994. Review of Computerized Profiling Version 1.0. Child Language Teaching and Therapy, 3:87–93.
Leonard, Laurence B. 1989. Review of the Pye Analysis of Language. Child Language Teaching and Therapy, 5:79–86.
Long, Steven H. 1991. Integrating microcomputer applications into speech and language assessment. Topics in Language Disorders, 11(2):1–17.
MacWhinney, Brain. 1992. The CHILDES Project: Tools for Analyzing Talk. Lawrence Erlbaum Associates, Hillsdale, NJ.
MacWhinney, Brian & Catherine Snow. 1985. The child language data exchange system. Journal of Child Language, 12:271–296.
Masterson, Julie J. 1995. Computer applications in the schools: What we can do—what we should do. Language, Speech, and Hearing Services in Schools, 26:211–212.
Miller, Jon F. & Thomas Klee. 1995. Computational approaches to the analysis of language impairment. In Paul Fletcher & Brian MacWhinney, editors, The Handbook of Child Language. Blackwell, Oxford, pages 545–572.
Ferry, Cecyle K. 1995. Review of Phonological Deviation Analysis by Computer (PDAC). Child Language Teaching and Therapy, 11:331–340.
Pollock, Karen E. 1988. Review of PEPPER: Programs to Examine Phonetic and Phonological Evaluation Records. ASHA, 30(8):57–58.
Pye, Clifton & David Ingram. 1988. Automating the analysis of child phonology. Clinical Linguistics & Phonetics, 2:115–137.
Rushakoff, Gary E. 1984. Clinical applications in communication disorders. In Arthur H. Schwartz, editor, The Handbook of Microcomputer Applications in Communication Disorders. College-Hill Press, San Diego, CA, pages 147–171.
Rushakoff, Gary E. & Arthur H. Schwartz. 1986. Clinical assessment software. In Michael L. Grossfeld & Cathleen A. Grossfeld, editors, Microcomputer Applications in Rehabilitation of Communication Disorders. Aspen, Rockville, MD, pages 1–24.
Schwartz, Arthur H. 1984. Microcomputer applications: Facts, functions, fads, and fallacies. Journal of Childhood Communication Disorders, 8:89–111.
Silverman, Franklin H. 1987. Microcomputers in Speech-Language Pathology and Audiology: A Primer. Prentice-Hall, Englewood Cliffs, NJ.
Sokolov, Jeffrey L. & Catherine E. Snow, editors, 1994. Handbook of Research in Language Development Using CHILDES. Lawrence Erlbaum Associates, Hillsdale, NJ.
Somers, H. 1978a. Computer analysis of speech therapists’ articulation tests. UMRCC Journal, 5(1):9–17 (University of Manchester Regional Computing Centre).
Somers, H. L. 1978b. Computerised Articulation Testing. M.A. thesis, Department of General Linguistics,
University of Manchester.

Somers, H. 1979. Using the computer to analyse articulation test data. *British Journal of Disorders of Communication*, 14:231–240.

Somers, H. 1998. Similarity metrics for aligning children’s articulation data. In *Proceedings of COLING-ACL ’98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montreal, Quebec, Canada, pages 1227–1232.

Stoel-Gammon, Carol & Paula Beckett Herrington. 1990. Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics & Phonetics*, 4:145–160.

Telage, Kal M. 1980. A computerized place-manner distinctive feature program for articulation analyses. *Journal of Speech and Hearing Disorders*, 45:481–494.

Van Demark, D. R. & Rosemary Tharp. 1973. A computer program for articulation tests. *Cleft Palate Journal*, 10:378–386.
