

SPAR'S Archivable Actual-Word Databases

Geoffrey Lindsey, Andrew Breen & Stephen Nevard

Phonetics & Linguistics. University College London

June 1987

Contents

- 0. Introduction
- 1. Reading material
 - 1.1 Actual words vs. nonsense
 - 1.2 Phonetic substance
 - 1.2.1 ACAWD
 - 1.2.2 APLAWD
 - 1.3 ASR vocabulary
 - 1.4 Sentences
- 2. Recording procedures
 - 2.1 Facilities and hardware
 - 2.2 Talker prompting
- 3. Notes
- 4. Tables and figures
 - Table 1. Rules for selecting phonetic items
 - Figure. 1. The prompting and recording arrangement
- 5 Appendices: the data
 - Appendix 1. ASR items
 - Appendix 2. The 5 sentences
 - Appendix 3. APLAWD 1- and 2-word items by phonotactic set
 - Appendix 4. ACAWfl 1- and 2-word items by phonotactic set

0. Introduction

This report documents the phonetic material, the talker prompting system, and the recording facilities, hardware and procedures used at University College to produce two databases for analysis within the SPAR (Speech Pattern Algorithmic Representation)

project. The recordings, of both speech pressure and laryngographic waveforms, are made digitally onto videotape and are archivable as Masscomp computer files in the Sparbase format.

1. Reading material

The data for the SPAR project is centred on two archivable databases, a Core database (ACAWD) and a shorter Priority List database (APLAWD), which is a subset of ACAWD.

1.1 Actual words vs. nonsense

This data is restricted to actual words (no nonsense-words), for the following reasons:

- (a) In order to constrain the size of the databases, it is helpful to exclude phoneme combinations which do not occur in English, such as word final short vowels (except /@/, /i/ and possibly /U/ in Received Pronunciation) or long vowels before the velar nasal /N/.<1>
- (b) Actual words are generally more likely to be given natural pronunciations than nonsense-words.
- (c) Nonsense phoneme combinations in transcription may be underspecified in crucial ways; for instance, VCV sequences can be stressed and syllabified in different ways. Actual-word data ensures more natural stress and syllabification.
- (d) Naive talkers as well as trained phoneticians are an integral part of the SPAR recording programme; such talkers cannot be expected to read non-orthographic transcriptions, and are likely to be confused by standard-orthographic representations of non- occurring words. Furthermore it is undesirable to use different core material for trained and untrained talkers.

1.2 Phonetic substance

The aim in both Core and Priority List databases is to represent in some sense a comprehensive account of the phonemes and phoneme-combinations of English. Given that the total number of combinations is impractically high, systematic restrictions have been placed on word-selection. These are outlined below.

1.2.1 ACAWD

ACAWD, the Archivable Core Actual-Word Database, contains 1691 one- or two-word items and five sentences. It aims to represent all occurring combinations of consonants and vowels, with the following restrictions:

- (a) The material attempts only to include CVC and VCV combinations (where C may be one consonant or a cluster or zero); each CVC or VCV may be embedded within a larger word, but no systematic attempt is made to represent longer combinations.
- (b) Only immediately adjacent slots are varied systematically for all occurring possibilities. The 'next door but one' slot is not systematically varied. Thus, for example, there are two sets of CVCs, which may be represented as CVC' and C'VC,

where the superscript denotes a slot which is not systematically varied. The CVC's in ACAWD include all occurring possible combinations of initial C and medial V, but no attempt is made in these items to include, for each CV, every occurring possible following C. Likewise, all occurring possible combinations of medial V and final C are included (C'VC), but these do not exhaust all the possible co-occurring initial Cs.

- (c) The CVC's include both a voiced and a voiceless final C for each preceding CV combination, in view of the substantial effects of final voicing on medial vowels.
- (d) The VCV items display both left-hand stress and right-hand stress; but left-stressed VCVs in ACAWD always syllabify the C-slot to the left, and right-stressed VCVs always syllabify the C-slot to the right, although the alternative syllabifications are possible (but rarer) in English. Thus ACAWD contains items like 'shap.ing but not 'may.pole (where ' represents stress on the following syllable and . represents a syllable boundary), and contains items like her.'soap but not dis.'eown.<2>
- (e) Where the C-slot is a cluster, the adjacent vowels are not varied exhaustively.

Table 1 lists the specific rules employed in selecting items to illustrate English phonotactics.

Note that the size of the database is reduced by the overlap between the different sets in the table above. For instance, the majority of the forms in the C'VC set are taken from the CVC' set.

The one- and two-word items in ACAWD are listed by phonotactic set in Appendix 4.

1.2.2 APLAWD

APLAWD (Archivable Priority List Actual-Word Database) is much smaller than ACAWD, comprising only 146 one- or two-word items and five sentences. It is intended for multi-talker and multi-repetition recordings.

APLAWD is a subset of ACAWD and makes a much more modest attempt at phonological comprehensiveness by exhibiting all taxonomic phonemes<3> in initial, medial and final positions (where these positions are occurring possibilities). As in ACAWD, the VCV sets include forms with syllabification and stress both to the left and to the right. In choosing items for APLAWD, priority has been given to the important ASR vocabulary and sentence vocabulary discussed in 1.3 and 1.4.

The full set of items in APLAWD (minus the five sentences) are listed by phonotactic set in Appendix 3.

1.3 ASR vocabulary

Both ACAWD and APLAWD contain a set of items considered important from the point of view of automatic speech recognition. These include the *alphabet*, the *digits 1-9* and *oh*, *zero* and *nought*, and several common and useful computing or office items such as *on/off*, *yes/no*, *plus*, *minus*, *file*, *letter*, *write* and *send*.

These items are listed in Table 1.

1.4 rive sentences

Both ACAWD and APLAWD contain five sentences which exhibit all taxonomic phonemes, though not in exhaustive environments. These sentences are composed entirely of words which occur in isolation in both databases.

One of the sentences consists entirely of vowels and liquids; another is composed entirely of phonologically voiced segments, although these include obstruents which are likely to be devoiced phonetically; one sentence is a wh-question one sentence is an imperative; one sentence is a mathematical equation.

These sentences are listed in Appendix 2.

2. Recording procedures

2.1 Facilities and hardware

The SPAR project recordings are made in an anechoic chamber in the Department of Phonetics and Linguistics at University College London.

The recordings are made with a Sony SL-F1UB recorder and a Sony PCM F1 digital processor. Both the speech pressure signal (Sp) and a laryngographic (Lx) signal are recorded.

Sp is recorded onto the left channel, using a Brüel & Kjaer half-inch pressure-sensitive condenser microphone (Type 4134) in conjunction with a Brüel & Kjaer microphone pre-amplifier (Type 2619). The signal is further amplified using a Brüel & Kjaer microphone amplifier (Type 2610) used with linear weighting 2Hz to 200KHz. The amplifier has the facility for high-pass filtering at 20Hz, which is not used in our recordings in order to preserve the low frequency phase characteristics of the speech pressure waveform.

The condenser microphone outputs negative voltage for positive pressure, which is the inverse of our conventional method of display. However, the PCM F1 processor inverts signals on input, so that the Sp signal is recorded with correct phase.

The Lx signal, from a portable laryngograph, is recorded onto the right channel. Because the PCM F1 inverts the signals it is necessary to pass Lx through an inverter before it is input to the processor, again ensuring correct phase on the recordings.

The talker is seated in an armchair with head restraint; the microphone head is positioned 15cm from the lips, 10cm from the mid-sagittal line and normal to a plane through and at the height of the talker's lips. Facing the talker is a monochrome monitor connected to the Masscomp computer, which prompts the talker with the data items.

Adjacent to the anechoic chamber is an annexe, where the Sp and Lx waveforms are displayed on an oscilloscope. Inside the anechoic chamber another oscilloscope showing Lx is used for correct placement of the laryngograph electrodes before recording.

After each recording, calibration signals are recorded on both channels. On the Sp (left) channel, a 1KHz tone at 93.6 dB SPL is recorded directly into the microphone, using a Brüel & Kjaer sound level calibrator (Type 4230). Further, a 20Hz 4:7 mark/space ratio pulse waveform is recorded onto both channels; for the Lx channel, a Laryngograph

dummy neck is used, and for the Sp channel a screwed-on calibration Jig is need to inject the waveform at the microphone preamplifier. (The jig consists of a capacitor with a value equivalent to that of the microphone.)

2.2 Talker prompting

The database items are presented to the talker on the monochrome monitor, in a different randomised order for each session. The one- and two-word items are presented at two second intervals, with a five second pause after every twenty-five items. The five sentences are presented at the end of the session, at five second intervals.

The prompting system is operated from a computer terminal in the annexe to the anechoic room, where the progress of the session is followed on a second monitor, in addition to the use of headphones.

APLAWD may be easily recorded in one sitting. A full ACAWD session lasts for just over an hour and the talker is given a break halfway through.

A diagram of the complete prompting and recording arrangement is shown in Fig 1. [Not included in this file]

3. Notes

1. Transcription here is in the machine readable phonetic notation recommended for the Alvey Speech Group.
2. Note that syllabification system adopted is that of John Wells, who posits attraction of surrounding Cs to the syllable of a stressed V, on the basis of such evidence as the shortening of a V before a voiceless C in the same syllable.
3. The notion of the phoneme used here is a traditional one, so that, for instance, the velar /N/ is treated as a phoneme on the basis of minimal pairs like sin and sing, although a more abstract analysis might treat [N] as an realisation of /n/.

4. Tables and figures

Table 1. The specific rules employed in selecting items to illustrate English phonotactics. Capital V represents a stressed V, lower case v an unstressed V.

COMBINATION	RULE	EXAMPLES
CVC'	final C voiced and voiceless; minimal pair preferred; /t/ and /d/ preferred; else obstruents preferred	pick, pig peck, peg pat, pad
C'VC	CVC' forms used if possible; else initial obstruent preferred	ship, Shep, chop
VC.v'	C'VC forms used if possible; final V: /-i/ preferred; else /-@/; else /-IN/	picky piggy, pity
v'.CV	CV portion from CVC' preferred; v': <i>her</i> before common nouns; <i>so</i> before adjectives; <i>she</i> before 3rd person verbs; <i>we</i> before bare verbs; <i>for</i> before proper nouns	her pig, her peg so poor she pays we give for Dick
(C)CCV	V: one front and one back, or one high and one low preferred	crude, creak
VCC	V: one front and one back, or one high and one low preferred	leaped opt, end
VCCC(C)	one only (final C is generally one of the alveolar inflectional endings); based on the VCC set where possible	opts, ends glimpsed
v'.(C)CCV	one only: where possible, made up of initial cluster form from (C)CCV set, preceded by a modifier as in the v'.CV	so crude
VCC.v'	one only (single-word forms rare, and two-word forms are marked due to English modifier-head order and the lack of modifiers ending in clusters); /-IN/ preferred after verbs; '-i' preferred after nouns	opting, inky

5. Appendices

Appendix 1.

Items in both ACAWD (Core database) and APLAWD (Priority List database) chosen for their usefulness in automatic speech recognition work.

Alphabet: A-Z

1-9

zero, oh, nought

ten, hundred, thousand, million

point

minus, plus, times, divided by, equals

calculate, double, percent

stop, start, end

load, save, run, move, speed

forward, reverse

up, down

right, left

of, at, from

true, false

yes, no

on, off

file, desk, letter, message, (write = right), read, send

Appendix 2.

The five sentences contained in both ACAWD (Core Database) and APLAWD (Priority List database):

1. George made the girl measure a good blue vase
2. Why are you early you owl?
3. Cathy hears a voice amongst SPAR's data
4. Be sure to fetch a file and send their's off to Hove
5. Six plus three equals nine