PROSPECTS FOR SPEECH TECHNOLOGY IN THE OCEANIA REGION

J. Bruce Millar

Computer Sciences Laboratory
Research School of Information Sciences and Engineering
Australian National University

ABSTRACT: The development of speech technology in the Oceania region is an issue for Australian speech scientists and technologists. In this paper we examine both the issues that govern the development of speech technology anywhere, the specific opportunities and inhibiting factors of the Oceania region, and the role that Australia, as the largest and most prosperous nation of the region, can have in the process. The necessary scientific resources required to establish both basic and more sophisticated speech technology are reviewed and mapped against the characteristics of the Oceania region. It is concluded that the most productive approach is likely to be one of creative partnership with the many island communities such that technology may be developed in a cost-effective and culturally sensitive manner.

1. INTRODUCTION

The development of new technology throughout the world is very uneven. Grand aspirations of universal benefit are typically voiced at its launch, then, as time progresses, the influence of other factors, such as individual greed within capitalism and individual lethargy within socialism, tend to erode the hope of benefit for those who are remote from the centre of power. By any reasonable measure a very large proportion of the residents of Oceania are geographically, economically and even linguistically remote from the power centres that drive speech technology. They suffer what we may call a three dimensional deficit of benefit.

Speech technology is developing at a rapid rate in highly developed regions of the world such as North America and Europe. It is providing access to information held by an increasingly wide range of public utilities and commercial companies via the nearly ubiquitous telephone handset. In this paper we will examine the potential for long-term deficit of such benefits that exists for the Oceania region and some measures that may be taken by the speech science and technology community to reduce the impact of such deficit on the residents of this region.

A full picture of the factors influencing the development of speech technology in the region will include the status of current language resources in the region and the prospects for their further development and their application to speech technology. These prospects will depend on relevant linguistic characteristics of the region, the level of technological development in the region with specific focus on the development of telecommunications, and, critically the relevance of access to available information to the communities of the region.

2. OVERVIEW OF THE REGION

The region is dominated in terms of land area, population, and economic prosperity by Australia (see Table 1). Many aspects of the region are characterised by the progressive move to independence from 19th century colonial powers. This pattern has resulted in a legacy of official languages made up of 81% English, 17% English-based pidgin, 1.5% French, and 0.6%

English/French. Populations range from 18.7 million in Australia to just 49 in the Pitcairn Islands. The economies range from the "western" economy of Australia with a gross domestic product per capita of 21,200 US dollars to the subsistence economies of Kiribati and Tuvalu whose domestic product per capita is just 800 US dollars. These data are derived from the USA Central Intelligence Agency (Central Intelligence Agency, 1999).

3. LINGUISTICS

The "official" languages of the Oceanic nations are strongly influenced by their colonial past which was dominated by Britain with smaller contributions from the USA and France. The official language will normally be the language of education and of formal administration. Its spoken form will be accented by the phonologies of the native languages of the speakers.

The "native" languages present a much more complex pattern arising from a history of tribal isolation and subsequent migration activity. Oceania has one of the lowest population to language ratios in the world. This makes the development of language resources linguistically complex and economically difficult. While there are certain anomalies, the Melanesian region extending off the north-eastern coast of Australia northerly to Papua New Guinea and easterly to Fiji, is characterised by many languages per island group (Pawley, 1995). In this region the number of speakers of a language can be of the order of 1000. In contrast, the regions of Micronesia, a northerly extension of Melanesia, and Polynesia, an easterly and southerly extension of Melanesia, are characterised by a single language per island group. Thus the indigenous languages of Polynesia and Micronesia are characterised by both a larger territorial range and population size than the those of Melanesia (Pawley, 1995). Throughout the region there are some 30 million people speaking approximately 1250 native languages.

The "contact" languages which represent the language of trade or other activities of forced contact between disparate language speakers often span much larger populations than "native" languages, and can rise in status to "national"

COUNTRY	Area sq.km	Pop'n 000s	Languages Spoken	GDP US$	Economy
Pitcairn Islands	47	0	English + Pitcairnese + Tahitian	n/a	Subsistence+fishing
Tokelau	10	1	English + Tokelauan	1000	Copra+Crafts
Niue	260	2	English + Niuean	1200	Subsistence
Nauru	21	11	English + Nauruan	10000	Phosphate(exhausted)
Tuvalu	26	11	English + Tuvaluan	800	Subsistence
Wallis and Futuna	274	15	French + Wallisian	2000	Subsistence
Palau	458	18	English + Palauan + 4 others	8800	Subsistence+fishing
Cook Islands	240	20	English + 4 other languages	4000	Supported by NZ
American Samoa	199	64	English + Samoan	2600	Fishing
Marshall Islands	181	65	English + Japanese + 2 others	1450	Copra+Fish+Crafts
Northern Marianas	477	69	English + Chamorro + Carolinian	9300	Tourist+Textiles
Kiribati	717	85	English + Gilbertese	800	Copra+Fish
Tonga	748	109	English + Tongan + 2 others	2100	Agric+Tourism
Micronesia	702	131	English + 4 other languages	1760	Farming & Fishing
Guam	541	152	English + Chamorro + Japanese	19,000	Military+Tourism
Vanuatu	14,760	189	English + French + 109 others	1300	Fishing,Finance,Tourism
New Caledonia	19,060	197	French + 38 other languages	11400	Nickel+Tourism
French Polynesia	4,167	242	French + Tahitian	10,800	Military+Tourism
Western Samoa	2,860	299	English + Samoan	2100	Agriculture
Solomon Islands	28,450	455	English + Pijin+ 120 other langua	2600	Ag+Fish+Forestry
Fiji	18,270	812	English + Fijian + Hind + 7 others	6700	Sugar+Tourism
New Zealand	282,680	3662	English + Maori	17000	Western
Papua New Guinea	462,840	4705	English + Tok Pisin + 714 others	2400	Agri+Minerals
Australia	7,686,850	18783	English + 234 other languages	21200	Western

Table 1: Areas, populations, languages and economies of the countries of the Oceania region.

languages as with Tok Pisin in Papua New Guinea. Some of these languages have developed from being second languages used mainly for trading purposes and living alongside native languages are labelled Pidgin languages) to being first languages for at least some of their speakers (and hence labelled Creole languages). Within the region there are six contact languages having speaker populations ranging between fifteen thousand and two million.

4. ECONOMICS AND CULTURE

When we view the geographical and linguistic characteristics of the region, the challenges for the development of speech technology are evident. One of the biggest of these challenges is to introduce this development in a culturally sensitive and economically viable manner. It is clear that speech technology developers, who already have a foothold in Australia, are looking to the bigger markets of Asia and beyond rather than to their Oceanic neighbours whose market is considered not viable. While this is a sound decision when viewed from the perspective of the shareholders in such speech technology companies, it suggests that a strong alternative view needs to be developed by those who are entrusted with public funding.

When considering the uptake of speech technology in a region it is imperative to examine the physical means, the financial cost and the motivation of the population of the region. Telecommunication development in the Oceanian region has developed slowly through four major network technologies. By the mid-1980s, the introduction of satellite telecommunication links of increasing sophistication and coverage, saw all nations with populations exceeding 30,000 hav-

ing some form of link (Ward, 1995). The 1990s have seen the advent of very high bandwidth fibre-optic cable but with typically longer runs which bypass many nations.

The incidence and rate of growth of fixed telephone lines in the region for the mid-1990s is found in data from the Telecommunication Development Bureau of the ITU (International Telecommunication Union, 1996). The incidence of telephone lines per 100 people divides the region into those countries with around 50 lines, those with around 20 lines, and those with 5 lines or less. The first group are those with essentially western economies, the second group are those where telephone use is strongly dominated by expatriate residents of colonial powers, and the third group represent a clearer picture of the status in independent developing nations of the region. It is in this third group that the challenge exists.

Telecommunication costs in the region have been analysed by Ward (1995). He indicates some of the factors that influence the very varied and sometimes asymmetric cost structures encountered. These structures are governed by the density of telecommunication traffic (influenced by factors of colonial past), official language affiliation, major trading partners, and local trade and travel. Ward concludes pessimistically that the island states of Oceania are likely to be excluded from information economy based on the cost of connection. It seems likely however that the cost of telecommunications for speech technology could be minimised by the location of call-centres which then link to information sources via the internet. It is also clear that the forms of voice data transmission are developing, with voice over internet protocols likely to rival existing voice circuits and more sophisticated ISDN connections. It seems likely that by concentration of "traffic" attracted by a relevant service, that the impact of telecommunication costs may be reduced. This more optimistic view is encouraged by reports from the region that indicate a very rapid uptake of internet access at main centres and that this is expected to spread to more remote islands and atolls where reliable telephone access exists (Early, 2000).

The question of what kinds of information would be most relevant for telephone users in the region does not as yet have a definitive answer, however scanning the major components of the national economies in table 1 will give an initial impression of application areas. It seems clear that information related to import and export of commodities, finance, and tourism is likely to feature strongly.

The "choice" of interface to such information, a graphical web-browser or voice, is worthy of consideration against the cultural and economic backdrop of the region. A web browser presupposes an investment in a personal computer and the skill to use it. While the use of the telephone may still be avoided on cultural grounds in some parts of the region, its acceptance must exceed that of a computer, and its relatively low cost and natural spoken language interface provides less impediment to its use.

5. INITIATIVES, OPPORTUNITIES AND INHIBITORS

We can now examine more closely the issues facing speech technology based on our review of the region. There are relatively few speech recognition or speaker verification systems in common use in the region. However there is strong speech-technology commercial activity in Australia and New Zealand at present. I am unaware of commercial installations of text-to-speech synthesis systems that are optimised to the local phonology despite a large research effort in Australia over many years. The major focus of speech research in Australia has been scientific and its has been driven to a large degree by academic rather than application goals. This has caused a focus on processes required to faithfully extract reliable information from speech data rather than on the data itself, its variability, its contamination, and where the limits of our knowledge demand that we adequately manage our ignorance (Millar, 1997).

The limiting factor in the development of speech recognition systems in Oceania is the need to acquire adequate data resources of local language patterns. Initial interest in the development of these resources has arisen in academic circles (e.g. Millar et al, 1990, 1994, 1997) where this development has been slow owing to labour intensive collection and annotation techniques. More recently a rapid expansion of speech data resources for telephone speech has occurred using the specific opportunities of commercial enterprises. One technique used has been to implement a rather simple speech recognition task using data models from a distant English speaking community, but to structure the task so that the perplexity of each step is limited in order to allow sufficient recognition accuracy to be obtained (Forsyth, 2000). If each recognised spoken token is also verified by a simple "yes/no" response, then the system can be used to collect reliable local data from which to develop local speech data resources. As these resources are expanded they can be used to build more accurate phone models which can support adequate accuracy with more difficult tasks. In this way the telephone speech applications can grow without the very heavy overheads of initial data model development. An alternative approach is to conduct an explicit data collection exercise and then integrate the data acquired with a distant model. Both kinds of process do, of course, produce data resources that are jealously guarded by their commercial developers as they have been gained by lengthy processes. Commercial enterprises do not wish to give competitors the opportunity to build sophisticated applications without the time delay inherent in their development.

The overall linguistic picture is complex and indicates that viable populations that speak the same language with a consistent phonology will be very few outside of Australia and New Zealand. However the region does have larger populations that speak accented forms of either English or French. One opportunity is for telephone speech technology to enter any population at a level of low perplexity recognition using external English or French models. This is initially tedious but has worked for applications in Australia. This development from low perplexity recognition is yet to be demonstrated in New Zealand. The penetration of such systems will clearly depend on the degree of accenting relative to the base models on which the externally developed systems are based. Studies that examine the "data increment" required to seed effective transformation of recognition performance, having a defined perplexity, from one accented form of English to another would shed light on the likelihood of success for this approach. Inhibitors to this approach may arise from a telecommunications cost perspective, and from any lack of relevance of available computer-based information.

Another initiative that could be examined is the analysis of the likelihood of telephone-based information access based on one of the "contact" languages such as "Tok Pisin" which has two million speakers. It seems clear that Tok Pisin has the semantic capacity to readily handle spoken information queries. Studies to reveal the realisation of Tok Pisin could help to assess the likelihood of success. It is clearly within the bounds of cross-speaker intelligibility but may strain the viability of current speech recognition techniques. It is significant for this approach that Tok Pisin has been systematised into a grammar and dictionary so the necessary information on which to build components in addition to phoneme models required for speech recognition systems are available (Mihalic, 1971).

The major inhibitors are perhaps threefold: the cost of development of systems relative to the economies of the beneficiaries; the availability of relevant information at a "telecost distance" (Ward, 1995) that is viable for the use of such systems; and cultural distance from the high-tech western world.

6. CURRENT DEVELOPMENTS AND ASPIRATIONS

Developments are currently focused in Australia and New Zealand but aspirations cover most of the region. There is a strong awareness amongst many leaders in the region that they must access the information highway of the world but the opportunities for the use of spoken language are not widely appreciated.

Overcoming the barriers of remote communities

The Community Teleservice Centre (CTSC) concept was developed for rural areas of Scandanavia in the 1980s and the derivative "Telecottage" concept emerged in the UK and Ireland in the early 1990s, and in Australia during the 1990s (Qvortrup, 1993; 1994). Qvortrup points to three barriers to the effective introduction of information services to remote areas: The service barrier due to services related to urban rather than

remote rural needs; the cost barrier due to low usage with respect to capital investment, and the qualification barrier due to the lack of skills in navigating through complex systems towards valuable target information. The CTSC concept can lower the cost barrier by concentrating traffic but it needs cultural awareness and appropriate technology to overcome the service and qualification barriers. Speech technology could overcome both of these if appropriate language resources are available.

Language Resources

Speech Technology is built on a foundation of language resources. These resources for speaker communities in developed countries have become the focus of intense effort in recent years. They comprise several facets, notably, a pronouncing dictionary in which the orthographic form of the language is linked to its phonemic form, a language model in which the way that dictionary entries (words) are typically strung together is defined in stochastic terms, and a set of phone models in which the range of acoustic realisations of each phoneme in a substantial range of contexts are also represented statistically. The development of these resources for a new speaker community can be prohibitively expensive unless they can be adapted from existing models for related speaker communities and/or be developed using a partnership between low-cost labour and high-tech facilities.

Developing Partnerships

It is very clear that commercial interests in speech technology cannot see any satisfactory return on investment in the region (beyond Australia and New Zealand). There is however a history of governmental largesse from these two countries to the smaller and less developed parts of the region. Strong economic development, political stability, and distinctive cultural attractions in the region are all important to the largely western economies of Australasia.

Partnerships that can build personal and institutional linkages, create economic and cultural benefits in the region, and exhibit cost-effectiveness in bringing speech technology to the region do appear to be feasible. The essential partnership is between language technology developed in high-cost western economies and language knowledge held in low-cost regional economies. It appears that many of the very labour-intensive parts of speech technology development could be performed by suitably educated members of target language communities. Tools for the building of pronouncing dictionaries, capturing adequate volumes of text for language modelling, and adequate recordings of speech for phone modelling can be provided at little added cost by the international speech technology community. The efficient use of these tools in a culturally sensitive manner can be most effectively managed by linguistically trained nationals.

It should also be noted that a major driver for such partnership is also the cultural survival of the region. Language is at the heart of culture and one impact of such partnerships will be to elevate the self-image of communities through the recognition that their own language can be a bridge to the high technology that is driving the information age.

A mechanism for regional partnerships

In 1991 the Coordinating Committee for Speech Databases and Assessment (COCOSDA) was established to unite speech scientists and speech technologists in creating best practice for creating resources for the development and evaluation of speech technologies. In 2000 this international consortium has been restructured to focus on key spoken language resource technologies across six major world regions, of which Oceania is one. The immediate task is to develop a network of experts in speech technology, in the linguistics of the region, and in the cultures of the region to examine the prospects for regional partnership.

7. CONCLUSIONS

Several challenges for the introduction of speech technology in countries of Oceania have been presented. As access to information is an essential factor for human development, there is impetus for assessing these challenges and then recommending the implementation of the appropriate technology to deliver that information. The modern portable multimedia personal computer is well adapted to provide a personal information interface. The ubiquitous presence of such interfaces and presentation software that is adaptable to all levels of human-computer familiarity would present an ideal human-information network. As this ideal is currently unattainable on economic grounds we must look at the role of the simpler interface offered by the telephone. Its implementation costs are low by comparison and the demands that it places on even unfamiliar users are far less.

It is therefore the claim of this paper that initiatives to examine closely the opportunities for the use of the telephone as an information interface using speech technology should be advanced in the Oceania region. The region presents a unique challenge to speech technology on account of its linguistic, geographic, and cultural composition.

ACKNOWLEDGMENTS

REFERENCES

Central Intelligence Agency, 1999. *Central Intelligence Agency - The World Factbook*, //www.odci.gov/cia/publications/factbook

Early, R. 2000. Personal Communication on 12 April 2000.

Forsyth, M., 2000, Personal Communication on 24 March 2000.

International Telecommunication Union, 1996, *Main Telephone Lines*, Oceania, //www.itu.int/ti/industryoverview/at_glance/deloc.htm

Mihalic, F., 1971. *The Jackaranda dictionary and grammar of Melanesian Pidgin*, Brisbane: Jackaranda Press.

Millar, J.B., Dermody, P., Harrington, J.M., Vonwiller, J. 1990. "A national spoken language database: concept, design, and implementation," *Proc. Int. Conf. Spoken Language Processing* (ICSLP-90), Kobe, Japan, pp.1281-1284.

Millar, J.B., Vonwiller, J.P., Harrington, J.M., Dermody, P.J. 1994. "The Australian National Database Of Spoken Language," *Proc. ICASSP-94*, Adelaide, I, 97-100.

Millar, J.B., Harrington, J.M., Vonwiller, J.P. 1997. "Spoken Language Resources for Australian Speech Technology," *J. Elec. Electro. Eng. Australia*, **17**, 13-23.

Millar, J.B. 1997. "Knowledge and Ignorance in Speech Processing," *Proc. Int. Conf. Speech Processing* (ICSP'97), Seoul, Korea, I, pp.21-27.

Pawley, A.K. 1995. "Language," in *The Pacific Islands: Environment and Society*, Bess, Honolulu, pp. 181-194.

Qvortrup, L. 1993. "Community Teleservice Centres and Rural Revival," *Proc. Telecottages '93*, Queensland, Australia, pp.69-82.

Qvortrup, L. 1994. "Community TeleService Centres: A means to social, cultural and economic development of rural communities and low-income urban settlements," *ITU World Telecommunication Development Conf.* (WTDC) Buenos Aires, 21 p.

Ward, R.G. 1995. "The Shape of Tele-Cost Worlds: the Pacific Islands Case," in A.D.Cliff, P.R.Gould, A.G.Hoare, and N.J.Thrift (eds.), *Diffusing Geography: Essays for Peter Haggett*, Blackwell, Oxford and Cambridge.