

Document Design for Learning English Pronunciation The swan-song of Gutenberg ?

Anthony Stenton
LAIRDIL and
Université Toulouse I Département des Langues,
Toulouse, France
anthony.stenton@univ-tlse1.fr

Saïd Tazi
LAAS-CNRS and
Université Toulouse I,
Toulouse, France
said.tazi@laas.fr

André Tricot
ERT 34 and Laboratoire Travail et Cognition
UMR 5551 CNRS
EPHE
Université de Toulouse le Mirail
andre.tricot@toulouse.iufm.fr

ABSTRACT One of the difficulties in English teaching lies in the correct use of primary and secondary lexical accents. French learners of English encounter particular difficulties in perceiving the stressed syllables and more particularly in encoding information on stress which is rarely necessary in French. They consequently have difficulties in producing primary and secondary accents when they speak. As the source of the problem seems related to attention, the dual coding of information on the word using sound and visual annotations is a potential solution. SWANS is an authoring system which enables the semi-automatic generation of multimedia documents in which accents are marked visually and sound is synchronized. This article presents SWANS, the hypotheses that underlie its development and the empirical evaluation of a model.

Introduction

In the context of what is referred to in France as the “neuro-cognitive pedagogical revolution”, ambitious claims have been made for the role of brain imagery in helping to elucidate learning problems. Teachers, it is declared, should behave more like doctors looking for innovative, individual solutions rather than imposing rigid, linear and cumulative stages in the ascent of an imposing mountain of knowledge. In the field of language learning, without resorting to brain scans, the very existence of such cognitive research has given a fresh impetus to applied research into new techniques for reading. In this article we argue that the plasticity of the computer environment can effectively mirror the plasticity of the human brain. The programme SWANS (Synchronised Web Authoring Notation System) developed by a group of 12 researchers working in four research laboratories in Toulouse, attempts to use synchronisation and enhanced typography to transform the experience of reading and listening. New exercises developed with SWANS tap into the brain’s adaptive capacities. The use of visual stimuli as a potential remedy for negligent auditory perception is possible because all the brain’s intelligences are connected. Consequently, the use of dynamic synchronised audiovisual events, more simply but, at times misleadingly, described as “karaoke for language learning”, may have implications for memorisation, comprehension and oral production. This ambitious challenge needs an adapted new technology to allow teachers not only to enhance the learning of pronunciation but also to allow teachers to practice new methods for visualising pronunciation to enhance teaching. The implementation of synchronisation and annotation technologies within an authoring system is the basis of the Swans. For the moment, SWANS generates web page documents and integrates audio and video materials synchronized with XML-based SMIL tags. These web pages are destined for the student. However, Swans is also designed as a toolbox for learning pronunciation. With the

final version teachers will be able to change even more typographical features and experiment new techniques for learning pronunciation and intonation. A further ambition is to open the system to other languages such as Spanish and German in the near future.

A notorious problem for francophone speakers of English

Students who study a foreign language must acquire knowledge about the stress patterns employed. For French students of English such learning is marked by a particular difficulty: as detecting lexical accent is less important in their own language even the perception of this accent in L2 is badly managed. Peperkamp who has studied stress contrasts across several languages speaks of the “stress deafness” in the French who, contrary to the speakers of more irregularly stressed languages, have little need to store information on generally predictable stress patterns in L1. Other empirical studies, see (Hawkins and Wartren 94), (Gupta and Mermelstein 82), (Dupoux et al. 97) suggest that francophones have:

- a) difficulties in perceiving stress patterns,
- b) difficulties in memorising them, which leads to
- c) problems of oral reproduction or production, which in turn has a negative effect on the comprehension of their discourse by English speakers. We have called this last problem the “tolerance threshold” and it is clearly related to individual experience and exposure to the L1 and L2 in question. Empirical research among an audience of English language teachers outside France suggests that, when fairly close together, even as few as 5 misplaced stress patterns (“neCESSary”) can lead to a serious breakdown of communication. Listeners simply stop making the effort to understand, though perfidiously they may continue to smile. Helping students to perceive, memorise and produce English stress patterns is thus a major challenge for university language teaching in France.

Typographic solutions

The idea of modified typography for helping with pronunciation has a long but erratic history. In 19th century France, the ‘Robertson method’ attracted considerable interest for several decades. More recently, Brazil attempted with cassettes, arrows and capital letters to teach both intonation and stress.

5.1

A rising tone is used in both versions of // GOOd EVening // etc.

2a // ↘ be**FO**RE // ↗ i introduce tonight’s **SPEA**ker // → there’s **ER** // ↘ **ONE** //
 ↗ important re**MIN**der //

3a // ↘ **NEXT** month’s // ↗ **MEE**ting // → will **BE** // → **OUR** // ↘ **ANNU**al **GENE**ral
 meeting //

4a // → **AND** er // ↗ on that **oCCA**sion // → we’re **HO**ping for // → a **GOOD** // → and
SPIrited // ↗ a**TTEN**dance //

Figure 1 showing annotations from Brazil 1994

In Figure 1, Brazil’s annotations have the advantage of reasonable visual clarity through using capitals and clear divisions of tone units with //. For students, however, it was agreed that the use of capitals might slow down reading speeds considerably and that colour might also be more effective for memorisation. It is important to note that the results have demonstrated that the following situations have been found to result in slower reading (Blustein 1999):

All upper-case print, italics, full-page bold, right justification with unusual spacing, white letters on a dark background. The transfer of Brazil’s paper-based annotation techniques to the computer screen is relatively simple but their synchronisation with sound requires new and adapted authoring tools.

Synchronising Solutions

The use of synchronized sound and text has been particularly timid in the field of language learning. This timidity is partly due to pedagogical doubts but must be linked principally to the absence of suitable tools for

synchronization. Although subtitlesⁱ have been used in the cinema since 1929, on television since 1938ⁱⁱ, and, in recent years, even in opera houses, and while karaokeⁱⁱⁱ has become a world-wide billion-dollar business for song learning, the tools necessary for synchronisation have been expensive, labour-intensive and, crucially, more adapted to audio-visual platforms than to the computer. Audio-visual material such as keyboards interfaced with video editing suites for encrusting on-screen characters were often judged too expensive for educational institutions. By contrast, most computer-based authoring systems either failed to address the question of synchronization or were prohibitively slow and cumbersome. Few teachers are willing to spend hours of painstaking trouble if the end product is limited to a few simple, synchronized sentences. For foreign language learning, the time factor explains the relative difficulties of such innovative tools as Audio Partner and Video Partner (Teleste) which, with characteristic Scandinavian sobriety, handled synchronization successfully as early as 1995. Closed-captions for video cassettes and, more recently and far more effectively, the DVD, have clearly increased awareness of synchronized subtitles as a potential learning aid but it is unlikely that synchronization will seriously interest teachers if authoring tools are not simple to learn and rapid to use.

SMIL and SWANS

An important breakthrough for permitting accurate fine-tuned synchronisation (measured in milliseconds) arrived in 1999 with the Synchronised Multimedia Integration Language (SMIL) as a recommendation from the W3C consortium. If the eye can recognise a word in L1 in an eighth of a second (Pinker) and the ear can recognise a word in a fifth of a second then our tools for exploiting such breathtaking accuracy must offer similar performances. “Only recently has temporal perception become a central issue again because cognitive processes cannot be understood without their temporal dynamics” (Poppel 1997). Speed reading techniques rely on reducing eye fatigue by flashing text in the middle of the screen thus avoiding the complex and time-consuming calculations needed to focus the eyes on the middle of the next words about to be read. By showing a block of blue colour behind synchronised text, the SWANS programme (2004), which uses SMIL language to generate synchronised web pages, permits a similar reduction in fatigue. The eyes are free to scan a highlighted line of text (not usually a sentence but tone units separated by pauses for breath) but are guided to the next line automatically in time with the playback of the sound. Contrary to traditional karaoke methods, this approach avoids the distracting hop from word to word which is intrusive as it troubles the field of vision as the eye scans the line backwards and forwards.

Our technique for treating stressed syllables in polysyllabic words using SMIL relied not on capitals but on increased size (12 replaced by 18) and blue letters. (Figure 2). The importance of colour for memorisation is still relatively unexplored but has to be examined in the context of computerized text. While colour cues have been found successful for localisation on hard copy (Blustein 1999) it cannot be assumed that the same effects hold true in an electronic environment. Research in the field of advertising suggests that colour stimulates interest 40% more than black and white (Kress and Van Leeuwen) whilst evidence from the study of the neurological condition called synesthesia (the concurrent response of two or more of the senses to the stimulation of one) reveals astonishingly individualistic and varied reactions to colour. Some findings suggest there are differences between gender in preferences for colours and reading speeds (Eysenck) but the fact that colour perception is fast, accurate, automatic, and effortless should not blind us to the fact that it is two-edged sword because improper use can create havoc with readability. Colour variation, just like Brazil’s use of capitals, can make grouping the letters into a readable word more difficult and hence slow down reading speeds.

Colour can be used to create cohesion in a text by uniting different elements with similar characteristics. Pronunciation rules, for example, such as the penultimate syllable stress of most words ending with “ic(s)” or final syllable stress for all words ending with two identical vowels (“kangaroo”, “trainee”) can be communicated or deduced by an appropriate use of colour and sound. In the context of the present project, however, the central hypothesis tested was one of dual coding. The focal point of explicit learning, the place of the primary and secondary accents, was encoded twice: visually and aurally. According to the theories of Paivio (1986), Meyer (2001) and Sweller (1999) such dual coding should lead to better learning for novice learners and have no effect or even a negative effect on learners who already know the place of these accents.



Figure 2. Showing text and annotations synchronized in an IE web page using the language Smil. The blue annotations can be shown or hidden as the user wishes.

Authoring synchronised and annotated texts with SWANS

In Swans, annotation is a set of typographical features added to a text to show rise and fall of the voice pitch while listening to the text. The act of annotation can be divided to three subtasks:

- The more difficult sub-task in annotating a text in Swans is to decide where to put on the annotation. For example, dictionaries often indicate where stressed syllables start but leave the user to deduce where they end. Syllables are like mountains - we can all 'see' the summit but identifying the frontiers of the foothills is not always easy.
- The second subtask consists of choosing the best typographical feature, that is adapted to the contextual substrings, according to the pedagogical goal.
- The last subtask consists of applying the annotation to the substring.

Deciding where to put the annotation and what kind of annotation to use depends on the intention of the 'author' (in this case a teacher/annotator). The on-going automation of annotation is based on two main ideas :

- the definition of a set of rules that represent the main features of English syllable stress patterns.
- the definition a set of metadata that a teacher can add to a text to represent his pedagogical intention.

In an ideal situation, if the text were tagged according to the author's intention, this knowledge would help in the generation of the annotation. This solution is not considered in this ongoing work.

For example, English words ending with the form 'tion' ["pronunciation"], have stress placed on the penultimate syllable. Such a rule will help by changing the typographical feature of the appropriate syllable. However, if the teacher specifies a different intention, the system will be programmed to annotate accordingly.

Synchronising text with audio allows simultaneous listening and reading. For the moment, this task is accomplished somewhat fastidiously via 'Magpie', a freeware from NCAM that generates time codes and produces output in SMIL format.

Swans is intended for teachers who are non-programmers. The aim is to help produce synchronised, annotated texts rapidly in a format appropriate for open and distance learning. In the medium term, it is also intended to extend the range of annotating techniques available for teachers and hence produce a wider variety of pronunciation exercises. The architecture of Figure 3 shows the different stages that lead to the generation of Web based documents containing annotated and synchronised multimedia materials destined for the student.

Architecture

Swans is composed of five modules presented in Figure 3. These modules involve the following functions :

- Importing text and media (video or audio) into the working environment.
- Segmenting the text into 'tone units' and tagging with XML codes. It should be noted that this stage is semi-automatic in order to leave the user free to choose appropriate units.
- Synchronising the text and sound. The freeware programme Magpie (Magpie 04) is currently used for this stage. After synchronising the programme checks the output code for coherence and puts the code into

SWANS format, that is to say a the script associated with each tone unit together with the start time and end time which is necessary for the precision of the karaoke. This stage involving the Swans format, a set of xml tags, allows the user to choose his own synchronising tools but more importantly will allow the future development of our own synchronisation system in order to accelerate the process even further.

- Annotating. The annotating tools available in Swans are deliberately manual and enable teachers to specify size, kerning and colour annotations of selected text simultaneously. Development underway of an expert system (based on Deschamps and Guierre) and a dictionary data base should accelerate the process by offering semi-automatic annotations which the user can validate or modify.
- - Generating ready to use web pages in XHTML+SMIL (W3C 02) This functionality, which provides the most dramatic increase in development speed, allows the user to visualise the document as it can be seen in Figure 2.

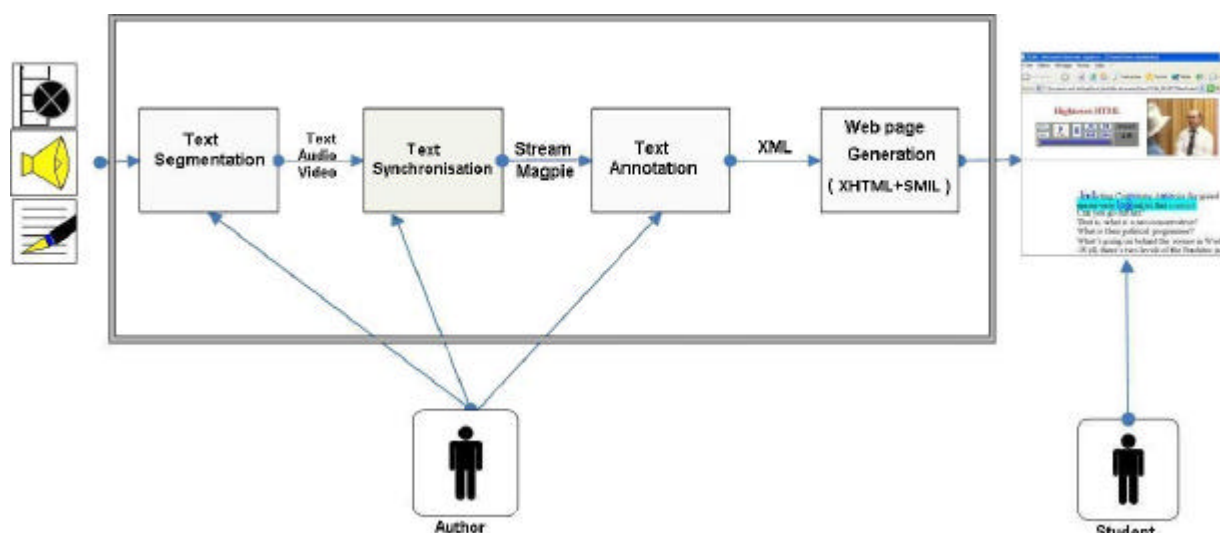


Figure 3. Showing the stages in the generation of a synchronised, annotated web page with SWANS

The evaluation of Swans

Informal testing of Swans began in 2004 with the generation of over 40 synchronized documents mainly from a bank of three-minute video news items. A considerable increase in speed of web page production was observed. The time taken to generate a web page of some 30 lines of annotated and synchronized text was cut from 2 hours to 10 minutes. Formal testing of Swans is due to start in June 2005 with teachers from a number of university language centres in Europe affiliated to CERCLES.

The evaluation of Dual Coding

Formal testing of the dual coding hypothesis started in 2004. Testing took place after presentations on screen using word lists which were annotated to show stress and linked to sound. Control groups were exposed to the same words with no annotation and the same sound. Students were requested to indicate the place of the tonic accent in two different written exercises and to record their own pronunciation of chosen words on the computer.

Our initial hypothesis, based on teaching practice, was that dual coding of primary stress would have a positive effect on novice students whether the feedback demanded was written or oral. This point of view was at the origin of the Swans authoring system. The corollary of the first hypothesis was that 'expert' students would show no improvement or a deterioration in performance called 'expert reversal effect'.

Participants

Testing included 64 second year undergraduates who had experienced relatively intensive training in pronunciation (hence the label 'experts'). A second group of 52 undergraduate students tested at the start of the 1st year were called the "intermediate" group as they had received no specific tuition on stress. Finally we were able to compare these results with those of "novices" - 50 pupils from a local high school (14 to 15 years

of age) who had never, or hardly ever, studied tonic accents (the teaching of clear speech in schools is rarely a high priority in France).

Results

Academic Level	Student production mode			C. oral
	Coding mode	A. written	B. written	
Experts (end of 2nd year students)	1. audio coding	9,26	8,69	7,55 *
	2. dual coding	7,45	6,50	5,19
Intermediate (start 1st year students)	1. audio coding	8,23	6,23	7,36
	2.. dual coding	8,37	7,11	7,17 *
Novices (High school)	1. audio coding	5,89	4,41	7,42
	2. dual coding	7,62	4,87	7,14

Table 1 : The effect of different coding methods on average performance according to academic level and to student production modes .¹

Globally the dual coding of the stressed syllable improves the performance of the novices and deteriorates that of the experts for the written exercise. By contrast, for the spoken exercise (where students recorded their own pronunciation on the computer), the dual coding has a negative or little effect on performance.

Average performance in the written exercises is between 5 and 9 out of 10. The second year university students have scores which are 0.5 points higher than those of 1st year students, who are themselves 2 points higher than the high school pupils. The dual coding improves the performance of the novices (by 1.1 points) and that of the intermediate students (by 0.5 points) while it has a negative effect (down 2 points) on that of the experts.

Discussion

The results of our study are encouraging because they show a positive effect for dual coding. The fact that this effect is not present in the oral exercise is perhaps due to insufficient data or an insufficient time lapse between presentation and testing.

Our results need to be completed. An experiment underway is now attempting to replicate these results with more complex and more convincing material from a didactic point of view (from word lists, we pass to discourse at the sentence level).

In time, we should be able to offer teachers an innovative tool for the generation of video web pages with annotated, synchronized scripts where tonic accents will be highlighted visually and whose effectiveness for language learning in terms of recognizing and producing appropriate stress will be attested through studies on novices at university and high school level.

Conclusion

Testing recognition of stress at the word level constitutes a prudent first step towards a much larger exploration of new intensive, and perhaps also extensive, reading techniques. Laboratory observation of students reading fairly long synchronised, annotated texts (10 minutes audio or video) suggests that improved concentration and reduced eye fatigue deserve serious analysis. Computer-based document design has slavishly copied the Gutenberg tradition since its inception. The arrival of SMIL in 1999 has brought the real potential plasticity of our electronic environments to the attention of developers.

¹ The asterisk * indicates groups with less than 10 individuals and thus statistically less significant.

Automatically generated textual annotations using colour, animation, flashing letters, changes in size and spacing offer a new way forward for multimodal research. Whether visual memory can really be enlisted to improve oral production or not, we are convinced that such experimentation will tell us more about students as unique individuals with unique ways of handling universal problems. The swan-song of Gutenberg has not arrived with a flash and a blare of trumpets but is part of a permanent quest to redefine the segmentation of, or the labels we give to, the experience of our senses. Like most evolutions, it will undoubtedly be slow and erratic and may well require several decades to defeat the resistance of ingrained habits. As Ong says, “freeing ourselves from typographical conditioning may be more difficult than we imagine” (Ong 82). Swans should be seen not as method for manipulating the minds of students but simply as an authoring tool for offering more choice in the variety of new reading techniques we can now place at their disposal.

Acknowledgments

This project is financed in the context of the CNRS TCAN programme 2003, with Anne Péchou, Nicole Décuré, Gail Taillefer, Antoine Toma, Christine Vaillant-Sirdey and Pascal Gaillard. Our special thanks to Nabil Kabbaj and Aryel Beck the principal programmers of Swans.

Bibliography

- Bickerton D., Stenton A., Temmerman M. (2001) Criteria for the evaluation of authoring tools in language education in ICT and Language Learning A European Perspective ed. Chambers, A. & Davies, G., Swets & Zeitlinger 20.
- Bickerton D., Ginet, A., Stenton, A., Temmerman M., Vaskari T., (1997) (Final Report of the RAPIDO Project, 91 pp., February 1997, University of Plymouth, UK (SOCRATES Project TM-LD-1995-1-GB-58)
- Blustein, W.J. 1999 Doctoral thesis, Department of Computer Science, University of Western Ontario, London, Ontario, Canada
- Brazil, D. 1994. Pronunciation for advanced learners of English. Cambridge University Press.
- Cazade, A. 1999. De l'usage des courbes sonores en apprentissage des langues. *Alsic*. Vol 2, Numéro 2, décembre : 3-32.
- Dechamps A. 1994, De l'écrit à l'oral et de l'oral à l'écrit, Orphyr
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, (1997) J., A destressing “deafness” in French? *Journal of Memory and Language*, n° 36, 1997, p. 406-421.
- Eysenck, H. J. (1941). A critical and experimental study of color preferences. *American Journal of Psychology*, 54, 385-394.
- Germain, A. & Ph. Martin. (2000). Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde. *Alsic*, vol.3, 1 : 71-86.
- Ginesy, M. (2000) Phonétique et phonologie de l'anglais, Ellipses
- Gupta, V., & Mermelstein, P. Effects of speaker accent on the performance of a speaker-independent, isolated-word recognizer. *Journal of the Acoustical Society of America*, n° 71, 1982, 1581-1587
- Guierre L. 1987, Règles et exercices de prononciation anglaise, Paris A. Colin-Longman
- Hawkins, S., & Warren, P., Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, n° 22, 1994, p. 493-511.
- Kress, G. and T. Van Leeuwen 1996 *Reading Images - The Grammar of Visual Design*. London: Routledge
- Levelt, W.J.M. 1989 *Speaking: From intention to articulation*. MIT Press.
- Levelt, W.J.M., et al (1991a) The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 98: 122-142.
- MAGPIE (2004) National Centre for Accessible Media, <http://ncam.wgbh.org/webaccess/magpie/> (consulted march 2005)
- Mayer, R. E., (1983) *Thinking, problem solving, cognition*, W. H. Freeman and company, New York
- Ong, W. (1982): *Orality and Literacy: The Technologizing of the Word*. London: Methuen
- Paivio, A. (1991b). Dual Coding Theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287.
- Péchou, A. & A. Stenton (2001) : Encadrer la médiation – le cas de la prononciation, *Asp 31/33 / Actes du Groupe d'études et de recherches en anglais de spécialité (Geras)*, Bordeaux Mars 2001 ISSN 1246-8185

Péchou, Anne & A. Stenton (2002): Encadrer la médiation – le cas de l'intonation,, Compréhension et Hypermédia, approches cognitives, communicationnelles et sémiotiques Albi

Peperkamp, S. 2001. Typologie des langues accentuelles : perspectives, développementales et données comparatives. Troisièmes journées internationales du GDR 1954 'Phonologie'. Nantes 30 mai, 1er juin 2001.

Pinker, S. (1984) Language Learnability and Language Development, Harvard University Press

Poppel Ernst 1997, A hierarchical model of temporal perception...in Trends of Cognitive Science n° 1, pp56-61

Schmitz, P. The SMIL 2.0 Timing and Synchronization model, Technical Report MSR-TR-2001-01 Microsoft Research, Redmond, WA, U.S.A. Available at: <http://www.w3.org/TR/smil20/smil-timing.html>. (Note : we are indebted to the expertise of Patrick Schmitz for the media control panel used in our authoring programme),

SMIL 2.0 Synchronized Multimedia Integration Language (SMIL 2.0) Specification W3C Working Draft 21 September 2000". W3C SYMM Working Group. Available at: <http://www.w3.org/TR/smil20/>.

Stenton, A.J. 1999 "The end of the cottage industry ? In-house hypermedia production in higher education." Proceedings of the 5th CERCLES International conference, Confédération Européenne des Centres de Langues de l'Enseignement Supérieur (Bergamo, Italie) Edited by D. Bickerton and Maurizio Gotti, Plymouth Cercles 1999

Stenton A., N. Kabbaj, S.Tazi (2003) : Remedial work for English Pronunciation with a smil(e) Europe-SMIL Conference, ENSAM, Paris

Sweller, J. (1988) Cognitive load during problem solving. Effects on learning, Cognitive Science, 12, 257-285

Sweller, J. (1999) Instructional Design in Technical areas, (Camberwell, Victoria,Australia: Australian Council for Educational research

Tricot, A., Lafontaine J. 2002 « Evaluer l'utilisation d'un outil multimédia et l'apprentissage » dans Le français dans le monde « Apprentissage des langues et technologies : usages en émergence »

ⁱ In the early days of film subtitling the main problem was to place the subtitles on the distribution copies, as the negative was usually in safe keeping in the country of origin. Norway, Sweden, Hungary and France quickly took the lead in developing techniques for subtitling films. However, the first country to subtitle seems to have been Denmark: in 1929 Al Jolson's *The Singing Fool* was shown in Copenhagen with subtitles. (Gottlieb, pp. 20-22)

ⁱⁱ On August 14, 1938, the BBC broadcast Arthur Robison's *Der Student von Prag* in a subtitled version. (This was probably also the first scheduled showing of a film in the history of television.)

ⁱⁱⁱ Karaoke is a word formed from putting two Japanese words together. "Kara" that comes from Karappo and means empty and "Oke", shortened from Okesutura meaning "orchestra". So Karaoke means "empty orchestra". The practise started in the 1970's in the city of Kobe. At present there are more than 100,000 "Karaoke boxes" or acoustically isolated rooms for singing without disturbing the neighbours in Japan. Daisuke Inoue, who is believed to have built the first karaoke machine in 1971 never patented the machine.
