Keller, E., & Bianchi, O. (2002). Virtual Historic Reconstruction with Speech Synthesis, in Braun, A. & Masthoff, H.R. (eds.) Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday (pp. 465-484). Stuttgart: Steiner. (Zeitschrift für Dialektologie und Linguistik. Beiheft 121.) ISBN 3-515-08094-5.

Virtual Historic Reconstruction with Speech Synthesis

Eric Keller[1] and Olivier Bianchi[2]
Laboratoire d'analyse informatique de la parole (LAIP)
[1]Informatique et méthodes mathématiques (IMM)
[2]Institut d'Archéologie et d'histoire ancienne (IASA)
Université de Lausanne, 1015 Lausanne, Switzerland
eric.keller@imm.unil.ch, www.unil.ch/imm – Synthèse

Introduction

Considerable improvements have been made in recent years in synthesising natural-sounding speech. Instead of robotic sound sequences that tax one's patience, the output of many contemporary systems is of acceptable and sometimes even pleasant quality, at least in certain contexts such as automatic information systems (e.g., computer-generated information, prices of goods, short announcements)[1]. It has thus been suggested that speech synthesis can now be used in wider contexts, such as in foreign language teaching, psycholinguistic experimentation and historic reconstruction (Keller & Zellner Keller, 2000; Keller, 2001). At this writing, it is only a matter of time before the first extinct language is recreated in speech synthesis[2]. Once a synthesised extinct language becomes available, virtual historical personalities can be recreated that are capable of a full range of speech and motion behaviour.

The complete synthetic reconstruction of the speech of an extinct language is bound to be both controversial and consequential. As far as we can identify, this type of reconstruction has never been undertaken before. Since it necessarily involves many uncertainties, the enterprise invites scientific challenge, not least of all because *representational reconstructions* generally have considerable scientific and popular impact, due to their ease of dissemination, their iconicity and their longevity (Moser, 2001). Before embarking on an enterprise of virtual reconstruction, it may thus be wise to pause and contemplate carefully the methodological procedures and the scientific implications of the enterprise, lest one be ready to be laughed off the podium.

The key issue is that of scientific authenticity. Given the impossibility of knowing every detail of a historical event or a historical linguistic state, some aspects of the reconstruction are necessarily inferred and others must be invented. In many cases, inferred components of the reconstruction outweigh historically attested components by a wide margin. Under such circumstances, how can a virtual reconstruction do justice to historical fact without inducing significant distortions? Consider a relatively easy project of the recreation of the discussions between Napoleon and French industrialists and military leaders, as well as French and European political personalities, that led up to Napoleon's disastrous invasion of Russia in 1812. What were the various voices like? Some authors mention Napoleon's strident and Corsican-accented voice[3], but what about the voices of the lesser-known interlocutors? And what was his spoken language like? To what extent can one extrapolate from dictations, bon mots and letters to Napoleon's interactive speech[4]? And what about the oral expression of attendants, generals and visitors? Since there are no easy

answers, some historians seek refuge in academic conservatism. They would argue that a historian's responsibility is to remain exceedingly cautious about unattested inferences, that virtual reconstruction should not even be attempted in the first place, and that it is better to remain silent on issues that cannot be settled on solid evidence.

Unfortunately, this is not really an option. Virtual reality reconstructions are inescapable, both for popular and scientific reasons. There is a considerable demand for documentaries and historic information by a public which is increasingly dependent on visual and spoken media. At the same time, 3D reconstruction and speech synthesis tools are becoming ever easier to use, which favours the creation of reconstructions for documentary television, museums, as well as university and secondary education. Beyond reconstructions destined for a popular audience, models and simulations created for scientific purposes test our scientific hypotheses with increasing realism. Speech synthesis used as a test bed for the integration of linguistic and historic knowledge is thus an inevitable future development, as we have argued previously (Keller & Zellner, 1997; Keller, 2001). Finally, it can be said that there is really not much use to historical and linguistic research, if its carefully obtained lessons cannot be conveyed to a wider audience, *with* the appropriate level of scientific nuance. A rational and systematic approach is thus required to guide historic reconstruction, an approach that must be formulated in sufficiently general terms to cover all aspects of the reconstruction, ranging from the historical events and utterances themselves, to the realism of the reconstructed locale, its objects, and the nature of the speech and voices employed.

We've chosen to explore this terrain with two small but representative examples of virtual reconstructions taken from the life of Caius Iulius Caesar. The methodological consideration of the various aspects of these two projects will lead to short discussions of each of these issues, the issue of reconstructing a classical Latin pronunciation and prosody will be considered in some detail, and a generalised framework for this type of simulation research will be formulated.

Methodological Considerations

The Scenes

The two scenes were chosen at different levels of reconstructive difficulty and historical significance. The first scene was to be as realistic and as 'doable' as possible under the circumstances, the second was to be more challenging and more significant historically. Both examples were chosen from Caesar's Bellum Gallicum.

In the first (titled the "Entrapment Scene", B.G. 3.17-3.19), we wished to have Caesar read aloud a text he had written or dictated and which contains his famous quotation 'People believe what they like to believe' (B.G. 3.18[5]). This scene involves a single, well-described historic personality, Caesar, reading in a relatively neutral and formal style, using the well-documented classical Latin of the Roman patrician class of the first century B.C. All aspects of the scene and the language employed are about as thoroughly researched and documented as can be expected for an inferred event occurring more than two thousand years ago. The scene was named the 'Entrapment Scene', because the famous 'believe'-sentence is but one of several reasons for an impatient and impetuous Gaulish attack upon a camp held by Sabinus, one of Caesar's lieutenants during the Breton, Norman, and Aquitanian campaigns of 56 B.C.[6]. Sabinus, who was fighting on the Norman front, in fact provoked the attack on his own fort by releasing a false Gaulish deserter who pretended that the Romans were out of food, weak, discouraged by rumours about the concurrent Breton campaign, and about to fold camp. The ensuing ill-prepared Gaulish onslaught was met by an organised Roman pincer

attack, and a complete victory over the attackers was achieved.

In the second example (the 'Divico Scene', B.G. 1.12-14), it was decided to recreate the remarkable event (58 B.C.) in which the elderly Helvetian representative Divico haughtily refused Caesar's request to stop the Helvetian's advance through Aeduen (Roman-allied Gaulish) territory, to leave hostages as a guarantee, to pay reparations to the Aeduens, and to turn around. As is well known, Caesar subsequently routed the Helvetians near Bibracte and sent the remainder of the Helvetian tribes back to what is now western Switzerland, thus paving the way for a rapid consolidation of the romanisation of that strategic area north of the Alps.

Despite its extreme brevity of three paragraphs, the scene is iconic of the entire Roman-Celtic conflict. It characterises the Gaulish presumptuous confidence at the outset of the Gallic wars, born as it was in the vast superiority of their numbers, their intimate knowledge of the terrain and their past transalpine military successes. And it opposes the equally brash answer of a Roman general perhaps half of Divico's age, inspired by Rome's relentless Mediterranean conquests, their impressive organisational and technical superiority, and Caesar's own irrepressible personal ambition. Two very divergent worlds were about to meet, and this brief exchange speaks volumes about the fate and profound disappointments that were about to befall the Celtic tribes of western Europe.

The technical difficulties of the scene are considerable. It involves the practically unknown Helvetian personality of Divico who is likely to have spoken through an interpreter whose command of classical Latin can be expected to have been less than perfect, as the Helvetians are known to have spoken a form of Gaulish. Furthermore, the speaking style, while formal, can be expected to have contained prosodic traces of interactivity and ill-disguised hostility. Finally, the original dialogue has to be reconstructed from reported speech. The Divico Scene thus involves many more reconstructive difficulties and unknowns than the Entrapment Scene.

The Reconstruction Fallacy and the Authenticity Index

In preparation for the re-enactments, a list of elements required for the two virtual scenes was drawn up, and linguistic, historic and archaeological rationales were sought for each (listed in Appendix 1 as formal structured computational variables destined for a 3D *scene graph*). A virtual re-enactment scene involving speech typically involved the following high-level information: (a) a global date and place for an initial subtitle, (b) a specific time and setting, needed for background scenery, objects and scene lighting, (c) one or several actors in period costume, (d) the script, or text in the original language and pronunciation, and (e) a translation into a modern language. This structure results in a well-organised set of historic or archaeological questions to ask concerning each element in preparation of the reconstruction proper. Space does not allow the presentation of the full historic and linguistic argumentation for each element in the two scenes. More information will be forthcoming in other articles and via our website[7]. The present report primarily addresses linguistic aspects of the project, as viewed from the perspective of the central theoretical problem of ensuring authenticity in virtual reconstruction. Some indication must thus be given concerning how authenticity was evaluated.

As suggested in the introduction, the problem of authenticity is pervasive and not only concerns linguistic issues. Let us illustrate the problem with respect to the place and time of the Entrapment Scene where, it will be recalled, Caesar re-reads the final version of volume 3 of *De Bello Gallico.* When and where did this event take place? Suetonius, based on earlier writings, says about 160 years after the event, "He [Caesar] wrote the first of these works

while crossing the Alps and returning to his army from Gallia Citerior, where he heard lawsuits" (Suetonius LVI). The Latin original makes it clear that "the first of these works" refers to the first volume (about 8'000 words), not all seven volumes written by Caesar (about 45'000 words)[8]. Indeed, the Commentaries' linguistic structure and the thousands of names and precise references in the entire work suggest that its final version was put together from a library of previous despatches and reports. This required staying in a single place for some time. Furthermore, the seventh volume, the last one written by Caesar himself, ends with a report that he was wintering in Bibracte ("He himself determines to winter at Bibracte." Caesar 7.90), while the eighth volume, written by Aulus Hirtius (Caesar's assistant), begins with Caesar's leaving Bibracte for Rome in January of 51 B.C. (Caesar-Hirtius 8.2). Also, the political situation at the beginning of the year 51 BCE necessitated a complete and heroic account of the war to prepare Caesar's return to Roman politics (Rat, 1964). It is thus reasonable to assume that Caesar reread his Commentaries while spending the first part of the winter 52-51 BCE in Bibracte, a well-established and central Gaulish hilltop oppidum. If so, where in or near Bibracte? In the habitation (i.e., an earthen-walled hut) of a Gaulish nobleman? In a stabilised Roman encampment? Or in a Gaulish meeting hall[9]? Depending on the answer, the backdrop to the reconstruction of the first scene will be the interior of a Gaulish habitation, that of Caesar's staff tent, or the interior of a Gaulish meeting hall. And obviously, there is no historic or archaeological evidence to establish the time of day.

This is quite typical of reconstruction efforts. Parts of the information are attested or inferrable, but many other details are not. Here, there is some general and reasonable indication of place and time, but it is not nearly precise enough for the stage setting of an animated and photo-realistic virtual re-enactment. The producer ends up weighing the probabilities or the stage effects of choosing one or the other solution, and ultimately makes a fairly arbitrary decision. This reposes the central question of authenticity: how much of the scene is directly attested, how much inferred from reasonable evidence, and how much is pure speculation? Traditional historical and archaeological writing uses *selection* and *argumentation* to stay on the acceptable side of scholarly argument: only elements that merit attention are selected for presentation, and argumentations concerning their significance are given where appropriate. This is a luxury not available to the reconstruction specialist whose creations must necessarily include arbitrarily chosen elements and whose leeway in providing scientific argumentations is greatly constrained.

It will not do to apply traditional means of differentiating attested and unattested elements in reconstructions either, because those are often excessively constraining, excessively complex, or impossible to apply in an animated virtual reconstruction context. For example, original and reconstrued segments of ancient buildings are often differentiated by the colour of the stonework, houses are reconstructed without rooves, since roof structure or materials are unattested, and some computer reconstructions provide options to display or hide poorly attested elements. No such option is open to the linguist attempting to recreate an extinct language, and an implementation of these devices would be very difficult with historic persons in an animated 3D context. In other words, the reconstruction specialist simply cannot avoid the *reconstruction fallacy*.

Since there is no way to avoid the problem, it must dealt with squarely. We propose to classify for *reconstructive weight* the principal elements making up a reconstruction scene. The resulting so-called "reconstructive weight index" (RWI) is composed of one of three levels of *verifiability* crossed by one of three levels of *significance.* The three verifiability levels are: (1) *possible*, i.e., judged possible given current sources, (2) *probable*, i.e., inferable as probable from current sources, and (3) *assured*, i.e., there is good support and no substantial counterevidence from current sources. The three significance levels are: (1) of

minor significance, (2) *significant*, (3) *highly significant* within overall intent of story or scene. The RWI of a given reconstruction element is thus the intersection between verifiability and significance. As an example of an element with very high RWI, Caesar's assassination was witnessed and reported by several persons, and it had major repercussions on Roman history. There is currently no reason to believe that this event is not historic. Both reliability and significance of the event can thus be classed 3. By contrast, the type of toga worn by Caesar at the Divico meeting can be inferred with a verifiability level of 2 (probable), and the relevance of the item would normally be assessed as 1 (trivial within the overall picture). That item thus carries a low RWI.

Table 1. Authenticity Scale as Applied to Caesar's Life Story

		Verifiability		
		1 possible (judged possible given current sources)	**2 probable (can be inferred as probable from current sources)**	**3 assured (there is no substantial counterevidence from current sources)**
Significance	**1 minor significance within overall intent of story**	Caesar stands up while reading his Commentaries	Caesar reread Commentaries in Bibracte during late autumn 52 BCE	Towards the end of his life, Caesar was balding
	2 significant within overall intent of story	Caesar lived in a staff tent during 52 BCE Bibracte wintering	Caesar was taller than his assistants	Caesar participated actively in battle
	3 great significance within overall intent of story	Towards the end of his life, Caesar believed in his own semi-divine status	Caesar was as a well-spoken, determined leader	Caesar was assassinated 44 BCE

Table 2: Authenticity Scale as Applied to Entrapment and Divico Scenes

		Verifiability		
		1 possible	**2 probable**	**3 assured**
Significance	**1 minor significance**	Optional furnishings during the meeting with Divico (i.e., useful, but not necessary objects)	Divico appearance: shirt of iron mail and long sword, tall, long blond hair, beard and mustache (typical appearance of noble Gaul)	Caesar appearance: military toga and tunic, tall, fair complexion, dark eyes, balding (attested characteristics for Caesar)
	2 significant	The language of Divico's interpreter: Gaulish-accented Latin	Language style at Divico meeting: fairly formal, cool (likely for a meeting between two military leaders whose troops have just fought a battle, and important for setting of Divico scene)	Caesar language: fluent use of Classical Latin (Caesar's excellent use of Latin is well-attested and is relevant to the Divico scene)
	3 great significance		Divico was probably surrounded by an armed escort.	Caesar's verbal reaction to Divico: clear and determined (Caesar reports it and he had a reputation for fast repartees--crucial in Divico scene)

In addition to defusing inevitable if justified criticisms of lack of authenticity in reconstructions, a systematic application of the RWI focuses scientific discourse on what really counts. For example, in judging reconstruction components, it is important to estimate

significance within the *overall intent* of the reconstruction effort. For example, Caesar's remaining seated in front of the temple of Venus while receiving the Senate (probably 45 B.C.) was judged very severely (Suetonius LXXVIII)[10]. In the context of the burgeoning signs of dictatorship and deification of the last years of Caesar's life, this apparent detail assumed great relevance and would be judged highly significant in a scene of Caesar's final years. By contrast, Caesar's being seated while re-reading his volumes is total conjecture, but the lack of significance of this detail renders an arbitrary reconstruction less objectionable (reliability 1, significance 1).

The scale is easy to learn and to apply. It can guide virtual reality designers in their work, and it can easily be employed for assessing the authenticity of a given part of a reconstruction effort. Within a standardised evaluation schema, it might even serve as the base of a 'quality label' for the entire reconstructive effort[11]. Although both concepts of verifiability and significance could be further refined on the basis of the strength of historical or archaeological evidence, we estimate that a simple scale with just three easily distinguished levels renders it more useful than a complex scale would. The purpose of the scale, after all, is to *encourage authenticity in reconstruction and to distinguish between reconstruction and fiction, not to assess the scientific reliability of a given archaeological or historic detail*. This tool thus encourages historic interpretation while providing an explicit assessment of both verifiability and significance of a given interpretation[12].

Speech Synthesis of an Extinct Language

Scientific efforts in support of historic reconstruction have concentrated by and large on events, appearance and function (see e.g., Biel, 1996; Coles, 1979; Ingersoll et al., 1977). If historic language use is recreated at all, it is generally pronounced by contemporary human actors, and most often, in modern translation. And since few modern actors are totally proficient in the phonology and prosody of the target language, the original language form is rarely reconstructed accurately; understandably so, because it is difficult to become a *fluent* and more importantly, a *precise and phonetically accurate* speaker of a language that has no native speaker population and whose native articulation cannot be heard or imitated on a daily basis[13]. Even taped readings by experts in classical Latin pronunciation such as Daitz (1984) show some weaknesses upon detailed analysis.

There is therefore some interest in employing speech synthesis for this purpose. If the phonetic and phonological details of an extinct language are adequately described, the reconstruction problem may well be sufficiently constrained to attempt a speech synthesis, with quality and reliability levels not too far removed from those obtained for modern languages[14]. Once such an attempt succeeds in its large outlines, experimentation with metric fit for well-documented poetic material can further constrain and provide external verification of the prosodic model. Finally, a systematic consideration of the verifiability and significance components of the RWI will keep unknown and unknowable linguistic and phonetic elements in perspective. Used in cautious and scientifically responsible manner, a reconstructed speech synthesis system can thus become a valuable experimental paradigm for testing our understanding of the fairly complex phonological and prosodic interactions of the target language.

Among long-extinct Indo-European languages, Latin, ancient Greek and Sanskrit appear to be good candidates for reconstructive efforts. The extant literature for all three languages is extensive, is well-researched and provides ample syntactic and morphological documentation. Closer to our project, fairly reliable information exists for the evolution of the Latin pronunciation of the patrician classes during the central centuries of Roman domination

in the Mediterranean and in Europe (roughly 100 B.C. to 300 A.D.). Much of the required phonological and metric information can be derived from the extensive writings of contemporary Roman authors, from Latin borrowings of Greek words into Latin and vice versa, from Latin into Greek, as well as from etymological evidence in later romance languages. In the present article, we shall pass lightly over evidence which is extensively documented in Allen (1978), Daitz (1984), Kent (1945), Niedermann (1906), Sturtevant (1940), Traina (1957), and which is used here but still unpublished (Bianchi, forthcoming).

Quite a different situation prevails for Gaulish, the form of continental celtic language spoken by Divico and his interpreter. A recent collection of all known or strongly suspected Gaulish words found on inscriptions, tablets and mentioned by Greek and Latin authors lists some 800 lexical items (Delamarre, 2001). The longest continental celtic texts, the hispano-celtic inscriptions on the Botorrita I and II tablets (found in 1971 and 1992 respectively) contain just a few hundred words (Lambert, 1997). Gaulish syntax and morphology is thus scantily attested, many details of the phonological system are still conjectural, and next to nothing is known about the phonetics of Gaulish (e.g., how strongly were plosives aspirated, or was there a perceptually relevant open/closed difference associated with the difference between short vowels and long vowels, etc.?). At current levels of knowledge, it might be possible to create a partial synthetic reconstruction of Gaulish with low verifiability levels. However such a reconstruction would remain largely untestable by external information, and would thus remain largely unproductive as a scientific enterprise.

Having said that, the extant phonological information on Gaulish does permit some hypotheses on interlanguage phenomena between Latin and Gaulish. For example, Eska (forthcoming) and Lambert (1997) list only two fricatives (the alveolar and velar fricatives /s/ and /x/) among the consonants of Gaulish, /f/ and /h/ being conspicuously absent (see discussion in Lambert (1997, p. 44). It is thus not unreasonable to suppose that Gaulish speakers would have experienced some pronunciation difficulty with those two common Latin sounds, perhaps substituting /p/ for /f/ and dropping the /h/ word-initially, where classical Latin apparently still maintained it[15]. Some commonly observed foreign-language behaviour can also be expected to apply, such as the fact that foreign-language speakers tend to shorten their sentences and that the cases of flexional languages tend to be confused (Grosjean, 1982). The latter hypothesis is particularly attractive for Gaulish speakers of Latin, who were likely to have shown confusion on the Latin accusative forms. Earlier celtic (Lepontic) and hispano-celtic attest the -am ending for the accusative singular in the a-stem nominal inflection, while Gaulish (which was in more direct linguistic interaction with Latin) shows -an, -em, -en and -im for the accusative singular (Latin corresponding form: -am) (Eska, forthcoming). This suggests considerable interlanguage effects, not only in Gaulish itself, but also in the Latin spoken by Gauls. An enlightened reconstruction of a Gaulish speaker of Latin, affected by Gaulish-Latin interlanguage effects, would thus seem quite possible.

In the next section, we shall outline how such reasoning can be used to drive a speech synthesis system for Latin within the three classic processing levels which are (a) graphemo-phonetic conversion, (2) prosodic, voice and style interpretation, and (3) speech signal generation.

Graphemo-Phonetic Conversion

After taking into account the various arguments and extensive, sometimes conflicting documentation that is often inspired by long-standing national traditions of Latin pronunciation, a phonological inventory and phonological rule set for classical Latin could be

determined with some certainty (Bianchi, forthcoming). The probable pronunciations were codified in an survey table, and the main phonological rules were identified and listed (Appendix II). Most of the phonetic and phonological information listed in the appendix turns out to be reasonably free of controversy. A long-standing discussion concerns the aspiration of plosives. Without going into much detail here, we interpret the evidence to say that the articulation of unvoiced plosives of early stages of Latin were relatively unaspirated, much like those of contemporary French or Dutch, and that a supplementary aspirated set of plosives entered into use with Greek loan words during the last century B.C. This apparently lead to the use of aspirated plosives in some native Latin words on the part of some members of the establishment[16] (for details of the argument, see Bianchi, forthcoming).

The main uncertainty that remains with respect to the phonological system concerns the exact degree of nasalisation of pre-final-/m/ vowels and degree of shortening or loss of final /m/. Classical Latin pronunciation in the first century B.C. appears to have shown some fluctuation, with greater degrees of nasalisation + final-/m/ loss attested for informal speech. The compromise adopted for the current project was to maintain such endings as vowel + /m/ combinations, but to shorten the final /m/ with respect to what its value would be in a modern European language like French, German or English.

Using these grapheme-phoneme correspondences and phonological rules, the Latin passages of the two scenes were translated into phonetic script by a Java programme and were manually corrected. The main problems that required human intervention at this stage involved the determination of vowel length. In many phonological contexts, vowel length is lexically determined and can only be identified through arduous verifications in the etymological dictionaries of various Romance languages. Currently, there exists no dictionary that lists vowel-length information separately (as against syllable length which is often noted), and the successive constitution of such a dictionary is one of the longer-term objectives of the current Latin reconstruction project. Overall, a surprisingly high verifiability level of 3 (assured) can be affirmed for this stage of the synthesis process.

Prosodic Information

Prosodic characteristics transmit important and delicate information concerning the speaker's discourse intent as well as his/her social and emotional status. This makes it particularly important to reconstruct prosodic information in believable fashion in the framework of a reconstruction project. However, both currently available speech synthesis technology and historical verifiability impose some decided limits on the modelling effort at this processing stage.

Expressed in the physical terminology favoured in the speech synthesis community, one may summarise prosodic processing as the determination of (a) duration values in *ms* for segments and syllables, (b) concurrent intonational (fundamental frequency, *F0*) values in *Hz*, (c) concurrent amplitudinal (*dB*) values, plus (d) supplementary acoustic indices relating to voice characteristics pertaining to stable speaker characteristics and the speaker's psychological and emotional status. Of these four types of information, the latter two (amplitudinal and supplementary acoustic information) cannot be satisfactorily manipulated by existing high-quality concatenative speech synthesis signal generation techniques, and they are largely irrelevant for the lower-quality formant-based systems. This presents a serious handicap for the synthesis systems of modern languages (Keller, 2001), or for that matter, for the recreation of an extinct language. Furthermore, the absence of a "living native speaker" or speech recordings for classical Latin introduces further limitations. For example, we have no means of assessing the precise (ms-measured) timing schema of classical Latin,

nor do we have much of an idea about melodic patterns (were Latin patterns more like those of modern-day Italian, more like those of contemporary French, or different altogether?[17]). The overall verifiability for prosodic processes is likely to vary between 1 (possible) and 2 (probable).

Should deficiencies at the prosodic level condemn the entire enterprise? Based on our preceding rationale, the answer is an emphatic 'no'. Along with historians and archaeologists accustomed to working with incomplete data, we would argue that there are different levels of incompleteness. We have seen that at current levels of understanding of Gaulish, the creation of a speech synthesis would amount to nothing more than building hypothesis upon hypothesis. In the case of classical Latin on the other hand, the gaps are inviting to be filled by a number of hypotheses of varying testability. As long as the effort remains transparent with respect to level of current verifiability, valid scientific progress can probably be made. Working rationales were thus developed for the timing and F0 of classical Latin, the two prosodic parameters that are typically manipulated in modern speech synthesis systems. Our rationale with respect to *timing* is presented first.

To begin, evidence from French poetry suggests that timing patterns remain surprisingly stable over time, as long as the *syntagmatic phonological structure* remains essentially unchanged. Phonological stability of this sort means an absence of loss or addition of entire segments. Changes in segmental quality alone (i.e., paradigmatic changes) are not seen as changing the syntagmatic phonological structure. For example, the change from rolled to uvular /r/ currently underway in rural parts of France and Quebec, or the merging of open and closed /a/ in modern Parisian French are not seen as affecting the syntagmatic phonological structure. However, historic changes involving the loss or addition of entire syllables or segments, especially those endowed with a certain prosodic prominence, are seen as having a major potential of bringing about prosodic change.

To illustrate the principle of syntagmatic phonological stability, we had our modern French speech synthesis system read French alexandrines from the 17th century (e.g., a portion of Corneille's Le Cid, written about 1637). Our system is trained on French prose as read by a proficient speaker of modern French. Yet it reads these alexandrines with largely appropriate rhythm, suggesting that the principal timing patterns of French have not changed much in three and a half centuries[18]. Also, a pilot project examining timing of the same sentences either embedded in prose or poetic contexts yielded no major timing differences at the syllabic level, suggesting that timing is governed by a single psycholinguistic process, no matter whether the speech material appears in poetic or prose contexts (Bianchi, 2001). This smoothes the logical path for proposing that the ultimate timing scheme should "work" (i.e., be a reasonable working hypothesis) for *any* prose and poetry from the entire period of classical Latin pronunciation during which the syntagmatic phonological system remained essentially stable in patrician speech. Also, if difficulties arise in making the timing scheme conform to documented or safely inferred metric patterns in prose and poetry passages, it will be taken as a suggestion that the proposed timing scheme is problematic. Furthermore, we have argued in our previous publications on the synthesis of French that much of the speech timing scheme is subject to universal articulatory constraints (e.g., Zellner Keller & Keller, in press). Although there are many differences to be found in European languages (e.g., with respect to final syllable lengthening), there is also a substantial degree of commonality in timing schemes, particularly among modern-day descendants of Latin. Furthermore, some dialects and sublanguages of the Italian peninsula, particularly those of Sardinia, are said to have preserved better than others certain aspects of the Latin paradigmatic phonological structure, at least with respect to their native vocabulary (i.e., excluding lexical borrowings from modern Italian and other languages). It may thus ultimately become possible to examine

the question of the reconstruction of Latin prosody through a more extensive study of various prosodic aspects of such Italic sublanguages.

However to get the ball rolling, we began with a simple working hypothesis that consisted of adapting to Latin the complex timing schemas worked out for French, and of empirically adjusting additional parameter controls for long vowels and stressed syllables in such a manner that the metrics seemed reasonable to the second author, an experienced reader of Latin prose and poetry. This portion of our model can therefore be considered to be at the level of an initial working hypothesis, to be examined and worked out in more detail in the course of future research projects.

Similarly, the *fundamental frequency* (F0) information represents one of the great unknowns of the project. Suetonius says that Caesar "is said to have delivered himself in a high-pitched voice with impassioned action and gestures, which were not without grace" (Suetonius LV). We also know that during the Gallic Wars, Caesar routinely addressed large assemblies of Roman soldiers (it can be assumed, through repeaters stationed throughout the assembly). A low-pitched monotone F0 for Caesar can thus be considered unlikely, but the exact intonational patterns are obviously unknown. Also, we know little about language-specific Latin F0 patterns for different types of sentences, particularly questions or commands. Fortunately, our pilot material involved primarily declarative sentences, and for these, we adopted a default strategy of providing a slight overall F0 declination, of placing a minor F0-"bump" for each syllable, of raising F0 more on stressed and lengthened vowels, and of giving extra prominence to semantically important elements[19]. But it is inescapable that hypothesis building in this domain remains largely untestable.

Voice and Speech Style Characteristics

Voice and speech style are a final speech component in need of consideration. A certain strength of voice may be deduced, even in monologue and dialogue situations, from Caesar's practice and talent in addressing large audiences. Latin authors provide some further information about Caesar's eloquence and speech style. Suetonius cites Caesar's contemporary Cicero as follows (LV): "Come now, what orator would you rank above him of those who have devoted themselves to nothing else? Who has cleverer or more frequent epigrams? Who is either more picturesque or more choice in diction?" Caesar's fluency and persuasiveness constrains the reconstructed speech patterns to be smooth and fluent, delivered in a commanding and convincing vocal style.

The problem with implementing these components may lie less with the historic reports than with the possibilities of current speech technology. What are the physical parameters associated with stable speaker characteristics? How is persuasiveness transmitted in the signal? Timing and F0 do capture some of these parameters, but other physical signal contributors are still unknown and are not yet open to manipulation[20]. Consequently, the best reconstructive strategy may be to choose, as much as possible, a diphone data base encoding a speaker whose stable characteristics match as best as possible the supposed voice quality of the target speaker, and to implement appropriate time and F0 manipulations[21].

With respect to voice and speech style characteristics, the second component of the Reconstructive Weight Index may well come into play: while the verifiability rating may be relatively low for voice and style (1, possible or 2, probable), the significance rating is likely to be low as well (1, minor significance). After all, not much harm is done to the overall scientific validity of the reconstructive effort if Caesar is presented with less than perfect voice and style characteristics.

Speech Signal Generation

Currently, concatenative diphone systems represent the main technology to support the final transformation of the phonetic and prosodic information into a speech signal. The freeware Mbrola project in particular permits relatively rapid implementation of experimental speech synthesis systems. For the current project, the various modern language data bases available for Mbrola processing were evaluated in terms of the following aspects: (1) a voice that is as clear, as easy-to-understand and as undistorted as possible, (2) a sound inventory that matches the classical Latin inventory as completely as possible, (3) the use of clear vowels, (4) a differentiation of open and closed variants for high and mid vowels, and (5) the availability of relatively unaspirated unvoiced plosives. While no modern language database satisfies all of these criteria completely, one database for Dutch (*nl2*, produced by Arthur Dirksen and Ludmila Menert of Fluency Speech Technology in Utrecht) comes surprisingly close. Speech produced by this database sounds pleasantly clear, and the Dutch phonological inventory largely matches that of classical Latin, the two sole gaps being the absence of short, open /ʊ/ and the use of a uvular /r/ instead of a rolled /r/. Also, the first French Mbrola database (*fr1*, produced by Thierry Dutoit of the Faculté Polytechnique de Mons, Belgium) satisfies several of these criteria. Despite the absence of clear distinctions of open and closed variants of high and mid vowels in modern French, the absence of the sound /h/ and the use of uvular instead of rolled /r/, reconstructions using this database are not unpleasant to listen to. The current pilot project was therefore executed with these two databases. We are currently examining the possibility of creating a diphone database that incorporates the full inferred phonetic inventory for classical Latin speech.

Evaluation and Multimedia Integration

Using these principles, a basic speech synthesis computer programme was created, and sound tracks for the first pilot scene were established. The evaluation and the multimedia integration of the sound material is currently underway, and the metric adequacy (see above), the reliability, as well as its ease of extension and maintenance of the system are being examined. Furthermore, sound tracks for the second passage, 3D personalities for Caesar and Divico, as well as historically and archeologically responsible scenery and objects are in the process of being created. More information about the visual aspects of this reconstruction will be forthcoming in future articles and on our website.

Conclusion

We have argued that the creation of a speech synthesis system for an amply documented extinct language like Latin affords considerable scientific benefits. In the context of total transparence concerning which portions are hypothetical, probable and attested, and a relative weighting of the significance of the various components, a computer recreation of the Latin speech patterns of the patrician classes of the classical period provides a valuable test bed for our understanding of syntactic, phonological and metric patterns of the language. Furthermore, the completed system will permit both specialists and laymen to experience Roman speech and culture in ever more realistic fashion.

The creation of speech synthesis systems not only summarises in most palpable fashion what we already know, it also shows us what we do *not* know (yet), and where our hypotheses are still thin. Only a truly complete synthesis system can ever sound like a human being, and the fact that no modern-language synthesis system truly sounds like a human in

extended stretches of speech shows that we still have much to learn about speech production, voice quality and speech style. Similarly, it is hoped that the blatant gaps in current knowledge about Latin timing and melodic structure, brought to the fore by the present attempts to build a speech synthesis system, will bring about a more systematic comparative analysis of such patterns in related modern languages and dialects. Ultimately, it is hoped that this will lead to further constraints on the system, and to an extension of the inferrable portions of Latin prosodic structure.

We wish to go one step further yet. In the long term, we estimate that the present speech synthesis system for classical Latin will become only one of many links within an extensive chain of computer simulations used for historical linguistics and the study of history and archaeology. For example, it is imaginable that all proposed and attested historical changes extending between modern-day Italian and supposed Indo-European root forms be stored in a master computer programme, and that various well-attested stages along the way be developed into full-fledged speech synthesis systems. Furthermore, the earlier and more hypothetical links along this chain could be interconnected with archaeological and historical simulations to form the basis of a wider understanding of the cultural and cognitive mould within which these language forms were used (for a partial and pre-computational initiative in that sense, see Renfrew, 1987). Also, later links in the reconstructive chain (such as late medieval or renaissance forms of various Italian dialects and sublanguages) could in conjunction with historic and archaeological evidence provide the basis of testable hypotheses concerning the complex cultural mould of the time.

As a final indication of the wide potential applications of computational simulation theory in this field, historical linguistics, history and archaeology have produced a bewildering array of theories and hypotheses in recent decades, based on massively increased amounts of empirical data obtained with the aid of often complex technical means such as genetic analysis, carbon dating, or in historical linguistics, by means of postulating formal constraints on possible diachronic change. If this enormous amount of information is not to suffocate in its own complexity, the undisputed or less problematic deductive steps must be computerised, leaving only permutable hypothetical information to be manipulated for scientific testing. In that manner, it will become much easier to link external archaeological and historic evidence to linguistic historic change in scientifically testable manner, and vice versa. Fields of enquiry can thus become linked for mutual verification that would not ordinarily be related by traditional means. In this fashion, the very palpable, "iconic" nature of speech and language reconstructions can become a valuable asset in making sense of our rich linguistic, historic and archaeological past.

Acknowledgements

We gratefully acknowledge stimulating conversations with François Zufferey (University of Lausanne), Alex Leukhard (emer. University of Geneva), Jean-Paul Guillaumet (Centre archéologique européen du Mont-Beuvray) and José Bernal (University of Lausanne), that shed much useful light on the arguments presented here. The research is supported by the Federal Office for Education and Science (BBW-OFES) of the Swiss Confederation.

References

ALLEN, W. Sidney (1978). Vox Latina: The Pronunciation of Classical Latin. Second Edition. Cambridge UK.

BIANCHI, Olivier (2001). La récitation de vers en français: quelques pistes d'analyse. (Verse Recitation in French: Some Possible Analyses). 10th Anniversary Meeting of LAIP, University of Lausanne, June 2001.
BIANCHI, Olivier (forthcoming). Vox rediuiua: synthèse de la parole et métrique latine. Doctoral thesis, Faculté des Lettres, University of Lausanne.
BIEL, Jörg (ed.) (1996). Experiment Hochdorf: Keltische Handwerkskunst wiederbelebt. Schriften des Keltenmuseums Hochdorf/Enz 1.
COLES, John M. (1979). Experimental Archeology. Chapter 1. London, pp. 1-48.
DAITZ, Stephen G. (1984). The Pronunciation and Reading of Classical Latin, a Practical Guide: Demonstration Texts and Practice Exercises. Guilford, CT.
DELAMARRE, Xavier (2001). Dictionnaire de la langue gauloise: une approche linguistique du vieux-celtique continental. Paris.
DODGE, Theodore Ayrault (1892). Caesar: A History of the Art of War among the Romans down to the end of the Roman Empire, with a Detailed Account of the Campaigns of Caius Julius Caesar. New York. Reprinted 1997.
ESKA, Joseph F. (forthcoming). Continental Celtic. In Roger D. WOODARD (ed.), Encyclopedia of the World's Ancient Languages, Cambridge, UK.
GROSJEAN, François (1982). Life with Two Languages: An Introduction to Bilingualism. Cambridge, Ma..
INGERSOLL, Daniel / YELLEN, John / MACDONALD, William (1977). Experimental Archeology. N.Y.
KELLER, Eric (2001). Future Directions of Research in Speech Synthesis. In Eric KELLER et al. (eds.), Improvements in Speech Synthesis. Chichester, UK. pp. 3-17.
KELLER, Eric / ZELLNER, Brigitte (eds.) (1997). Études des Lettres, vol 3.: Les défis actuels en synthèse de la parole. Lausanne.
KELLER, Eric / ZELLNER-KELLER, Brigitte (2000). New Uses for Speech Synthesis, The Phonetician, 81, 35-40.
KENT, Roland G. (1945). The Sounds of Latin: A Descriptive and Historical Phonology. Baltimore: Linguistic Society of America. (Originally published 1932, reprinted 1979, Milwood, NJ).
LAMBERT, Pierre-Yves (1997). La langue gauloise. Paris.
MOSER, Stephanie (2001). Archaeological conventions: The visual conventions for constructing knowledge about the past. In Ian HODDER (ed.), Archaeological Theory Today. Cambridge, UK. pp. 262-283.
NIEDERMANN, Max (1906). Précis de phonétique historique du latin. Paris. (Reprinted 2000).
RAT, Maurice (1964). Préface. In César, la guerre des gaules. Paris.
RENFREW, Colin (1987). Archeology and Language. Reprinted 1998, London.
SANTI, Serge, GUAÏTELLA, Isabelle, & LAGRUE, Benoît (2001). La représentation physique d'une personne à partir de sa voix: méthodologie et résultats préliminaires. In Christian CAVÉ et al. (eds.), Oralité et gestualité: Interactions et comportements multimodaux dans la communication. Actes du colloque ORAGE 2001, Aix en Provence. pp. 435-441.
STURTEVANT, Edgar H. (1940). The Pronunciation of Greek and Latin: The Sounds and Accents. Philadelphia. (Originally published 1920; reprinted 1975).
TRAINA, Alfonso (1957). L'alphabeto e la pronunzia del latino. Bologna: Pàtron. (Reprinted 1973).
WILLEMS, W.J.H. (1993). Archäologische Denkmalpflege und Forschung in den Niederlanden. Archäologische Denkmalpflege und Forschung. Hrsg. vom Thüringischen Landesamt für Archäologische Denkmalpflege durch Sigrid DUŠEK (Weimar). pp. 22-27.
ZELLNER KELLER, Brigitte (work in progress). Speech synthesis and speakers' prosodic styles.
ZELLNER KELLER, Brigitte, & KELLER, Eric (in press). The chaotic nature of speech rhythm. In Ph. Delclocque and V.M. Holland, Integrating Speech Technology in Language Learning. Swets & Zeitlinger.

Sources

CAESAR, Iulius Caius. De Bello Gallico. Translators W. A. MCDEVITTE & W. S. BOHN. 1st Edition. New York. Harper & Brothers. 1869. Harper's New Classical Library. OCLC: 25172949 (also multiple Internet locations).
PLUTARQUE. Vies parallèles I. Vie de César. Traduction J. Alexis PIERRON. Paris. 1995.
SUETONIUS. De Vita Caesarum, Divus Iulius. Translation J. C. ROLFE, Cambridge, MA and London, 1920. (Internet: Ancient History Sourcebook).

Appendix I: Partial scene graph attribute list for Caesar re-reading aloud passage from Book 3.18

Node	Object	Attribute	Evidence
Setting	place.geographical	wintering_Bibracte	inferred, good evidence, see text
	*.locale	Gaulish_meeting_hall	inferred, weak evidence, see text
	*.locale.structure	timber structure	inferred, weak evidence
Actor	person[1].name	Caesar	known
	*.body.form	sturdy	inferred, Suetonius 45, Caesar statues
	*.body.size	tall	reported, Suetonius 45
	*.body.bearing	noble	reported, Cicero, cited in Fuller p. 53
	*.head.form	inverse pear-shaped head	inferred, Turin statue (only preserved statue probably created during Caesar's lifetime)
	*.head.hair.form	wavy	inferred, Turin and other Caesar statues
	*.head.hair.hairline	receding	inferred, Turin statue, comments of contemporaries
	*.head.eyes.colour	black	reported, Suetonius 45
	*.head.hair.colour	black	inferred, weak evidence, relationship to eye colour
	*.skin.colour	fair_compexion	reported, Suetonius 45
	*.action	reading_from_papyrus	inferred
	*.speech.language	classical_Latin	inferred, Commentaries, etc.
	*.speech.pronunciation	classical_Latin	inferred, Allen, Bianchi, etc.
	*.speech.speechStyle	reading_aloud	inferred
	*.speech.dialect	aristocratic_instructed	reported, Suetonius 55
	*.speech.idiolect.pitch	high-pitched	reported, Suetonius 55
	*.speech.idiolect.rhythm	fluent	reported, Suetonius 55
	*.dress	aristocratic_tunic	inferred, Caesar statues
	*.dress.colour	white, brown	inferred
Props	furniture.container.maps	size, form?	inferred, typical Roman furniture
	*.container.papyrus	size, form?	inferred, typical Roman furniture
	*.seat	size, form?	inferred, typical Roman furniture
	*.table	size, form?	inferred, typical Roman furniture
	wall_hanging.map_of_Gaul	unrolled parchment	inferred, Peuter's map

Annex II: Grapheme – Phoneme Correspondences for Classical Latin and Relevant Contact Languages in first century B.C.

Greek graph	Greek pron	Classical Latin graph	Early Latin pron	Classical Latin pron	Vulgar Latin pron	LAIP Latin phon alph	Gaulish graph (Latin script)	probable Gaulish pron
Π	p	P	p	p	p	p	P	p
Φ	p^h	PH	p	p^h	p	P		
Β	b	B	b	b	b	b	B	b (soft, sometimes confused with /m, w/)
Τ	t	T	t	t	t	t	T	t
Θ	t^h	TH	t	t^h	t	T	τ (tau Gallicum)	various: ts, ds, st
Δ	d	D	d	d	d	d	D	d
Κ	k	C	k	k	k	k	C	k
Κ	k	K (rare)	k	k		k	(sometimes K, G)	k
Χ	k^h	CH	k	k^h	k	K	X (sometimes)	χ (before s, t)
		QV	k^w	k^w	k^w	kw	Q	q^w (rare, archaic)
Ξ	ks	X	ks	ks	ks	ks	X	ks
Γ	g	G	g	g	g	g	G	g
		GV	g^w	g^w	g^w	gw		
Ν	n	N	n	n	n	n	N	n (ŋ before velars)
Μ	m	M	m	m	m	m	M	m
Ρ	r	R	r	r	r	r	R	r (sometimes r^h)
Λ	l	L	l	l	l	l	L	l
		F	f	f	f	f		
Σ	s	S	s	s	s	s	S	s
Ζ	z	Z (borr.)		z	z	z		
		H (initial)	h	h	h	h		
		H (intervocalic)	ʔ	ʔ	ʔ	q		
Ι	iː	I (long)	iː	iː	iː	iː	I	iː
Ι	ɪ	I (short)	ɪ	ɪ	ɪ	I	I	i
		I (unaccented, esp. pre-labial)	ɨ	ɨ	ɨ	*		
Ι	i	I (S-V)	j	j	j	j	I	j
Υ	y	Y (borr.)	u	y	y	y		
		V (long)	uː	uː	uː	u	EU	uː
		V (short)	ʊ	ʊ	ʊ	U		u
		V (S-V)	w	w	w	w	U (late: F)	w, β?
Η	ɛː	E (long)	eː	eː	eː	e	E	eː
Ε	e	E (short)	ɛ	ɛ	ɛ	E	E	e
Ω	ɔː	O (long)	oː	oː	oː	o	O	oː
Ο	o	O (short)	ɔ	ɔ	ɔ	O	O	o
Α	aː	A (long)	aː	aː	aː	a	A	aː
Α	a	A (short)	a	a	a	A	A	a

Notes:

- This table is organised according to *graphemic* principles. Some Latin graphemes convert to more than one phonetic segment (e.g., "X" → /ks/).
- The *Latin* components of this table are based on Allen 1978, revised 1988, and O. Bianchi (forthcoming).
- The *Greek* components of the table represent the probable pronunciation at the time considered (1st century B.C.).
- The *Gaulish* components of the table are based, in this order, on information from Eska (forthcoming), Lambert (1997), and from Christopher Gwinn's (2000) web page on Gaulish (members.nbci.com/gaulishweb/phonol.htm). The three sources do not always agree. In addition to the Latin script predominantly found in late inscriptions, the Gauls used other local scripts such as the Greek and the Etruscan scripts, with each script "filtering" the Gaulish phonetic output in its own manner. As a general rule, phonetic interpretations of Gaulish graphemes should be considered as subject to much greater remaining uncertainty than those of classical Latin.
- The *LAIP phonetic alphabet* is a computer-readable codification of the Latin phonetic transcriptions, constructed in analogy with the Laboratory's computer-readable phonetic scripts developed for French, German and English. The script only uses letters found in the 128 lower ANSI alphabet which are implemented in uniform fashion on the Windows, UNIX and Macintosh platforms. Also, the present alphabet for classical Latin is constrained by the one-letter-one-sound standard, which considerably facilitates and accelerates computer processing of phonetic material.
- In accordance with common use, small case is used in these Notes for Latin spellings.
- Pronunciations are given in IPA-93. The small-case for /r/ is significant, it designates a rolled /r/.
- In the patrician speech of the classical period, Greek borrowings written with "Φ", "Θ" and "X" received aspiration /*PC*^{h}/. For the rest, the unvoiced plosives of the classical period are likely to have been, for the most part, similar to French pre-vocalic unvoiced plosives (see text).
- The "qu" /k^{w}/ and "gu" /g^{w}/ were probably rapid transitions, more like /k^{w}/ as in "squat" and /g^{w}/ in "Guatemala" than the /kw/ and /gw/ of slow English pronunciations of "qualm" and "Guam".
- The higher long vowels /i:, e:, u:, o:/ tended to be pronounced with closed articulatory variants, while the high and mid short vowels /ɪ, ɛ, ʊ, ɔ/ tended to be pronounced with more open articulatory variants. Consequently, the length marks are suppressed in our phonemic transcriptions.
- The pronunciation of /f/ and /s/ was probably similar in all environments during the classic period (i.e., no voicing, so "Caesar" was /ka^{j}*s*ar/).

Some Latin pronunciation and assimilation rules:

- Double consonants are pronounced as prolonged ("ille" /ɪlːɛ/, "agger" /agːɛr/, "fissus" /fɪsːʊs/.
- *Plosive-devoicing rule:* voiced plosives preceding unvoiced consonants are devoiced and involve a lenis-fortis change "plebs" → /plɛps/, as witness the presence of alternate spellings ("pleps") and full assimilations ("obfero" → "offero"). Thus voiced plosives preceding unvoiced sounds are rewritten as unvoiced plosives in the phonetic transcription.
- *Exceptional assimilation rule:* certain sound sequences, such as "abf", "obf", "obm", "obg", "obc", etc. are most frequently assimilated: "abferro" → /af:ɛr:o:/.
- *G-nasalisation rule:* g→ŋ /_n, e.g. "ignes" /ɪŋneːs/, "cognatus" /kɔŋnaːtʊs/. Attested in alternate spellings, e.g., "ingnes". Word-initially, /g/ is sometimes preserved: "Gnaeus" /gnajʊs/.
- *Nasal-assimilation-nasalisation rule:* n→N/_$C_{plos\ vel}$ ("lingua" /lɪŋg^{w}a/) and m→N/_$C_{plos\ vel}$ ("tam grauis" /taŋgrawɪs/), as well as n→m/_$C_{plos\ lab}$ ("in pace" /ɪmpa^{i}kɛ/) and m→n/_$C_{plos\ alv}$ ("tam durum" /tandu:rʊm/).
- *Final-m-nasalisation:* "dominum" /dɔmɪnõ/, "scriptum est" /skrɪptõst/. The final syllable counts as heavy.
- *Liquid-assimilation:* r→l/_l, e.g. "perlego" /pɛlːeːgoː/.
- *H-weakening:* h → ?/V_V, e.g., "nihil" /ni?il/ or /niːl/, "mihi" /mi?i/ or /miː/.
- *Diphthongisation:* graphemic vowel groups having /i/, /e/ and /u/ in second place, such as "ai", "ae" and "au", are interpreted as corresponding diphthongs, e.g. /a^{j}/ and /a^{w}/.

Footnotes

[1] The French version of our laboratory's speech synthesis system "SpeechMill" is used for announcements for an experimental radio station (www.klemradio.com). See also the online version at www.unil.ch/servlets/imm/SyntheseServlet.

[2] A speech synthesis for classical Latin is currently under construction in association with our laboratory (www.unil.ch/imm/docs/LAIP/LAIPTTS_old_lang.htm).

[3] http://ecole.cravanche.free.fr/napoleon.htm.

[4] "Madame de Rémusat: From Memoirs. Bonaparte dictated with great ease. He never wrote anything with his own hand. His handwriting was bad and as illegible to himself as to others; and his spelling was very defective. He utterly lacked patience to do anything whatever with his own hands. The extreme activity of his mind and the habitual prompt obedience rendered to him prevented him from practicing any occupation in which the mind must necessarily wait for the action of the body. Those who wrote from his dictation--first Monsieur Bourrienne, then Monsieur Maret, and Méneval, his private secretary--had made a shorthand for themselves in order that their pens might travel as fast as his thoughts. He dictated while walking to and fro in his cabinet. When he grew angry he would use violent imprecations, which were suppressed in writing and which had, at least, the advantage of giving the writer time to catch up with him. He never repeated anything that he had once said, even if it had not been heard; and this was very hard on the poor secretary, for Bonaparte remembered accurately what he had said and detected every omission. . . ." (quoted from the Internet Modern History Sourcebook, http://www.fordham.edu/halsall/mod/remusat-napoleon.html).

[5] The Perseus online edition numbers this passage as 3.16 (www.perseus.tufts.edu).

[6] Dodge (1892) places the site of the battle at the Camp du Chastellier, situated four miles east of modern Avranches, Normandy (pp. 139-143).

[7] See in particular www.unil.ch/imm/docs/LAIP/LAIPTTS_old_lang.htm and www.unil.ch/imm/docs/LAIP/LAIPTTS_L_SpeechMill_dl.htm.

[8] Suetonius LVI " quorum librorum primos in transitu Alpium, cum ex citeriore Gallia conuentibus peractis ad exercitum rediret", "[He wrote] the first of these works while crossing the Alps and returning to his army from Gallia Citerior, where he heard lawsuits." (Rolfe translation).

[9] That may well be the strongest hypothesis, according José Bernal, of the University of Lausanne's *Sciences de l'antiquité* section who is field supervisor of excavations of the Roman villas found in the "Equestrian Park" area of the Bibracte hilltop settlement on Mont-Beuvray (Burgundy, France). The solid-stone Roman structures found there appear to postdate Caesar's 52-51 B.C. passage at Bibracte. At the same time, the remains of a number of large Gaulish halls based on an elaborate timber construction have been identified at Bibracte. Such a hall might have done excellent duty as headquarters for Caesar's staff during his extended stay of several months.

[10] Both Suetonius (LXXVIII) and Plutarch (60) emphasize the severe hatred that Caesar's remaining seated engendered in the Roman population. Plutarch (60) reports that Caesar offered the somewhat lame excuse that standing up exposed him to epileptic seizures.

[11] This may turn out to be more difficult than it would seem at first, since many virtual objects contain both attested and conjectured elements. If one were to take Caesar's physical appearance as an example, Suetonius comments on his dark eyes, but nowhere are we informed about the colour of his hair. Negative evidence (the fact that no one mentions it) and a probable, genetic (melaline-based) correspondence between darker colours for eyes and hair suggest a darkish hair tint, but at this point, this is largely conjecture. In the context of a systematic and global evaluation of authenticity, every detail of every scene would therefore have to be judged individually. This renders the authenticity scale more useful for the designer of virtual reconstructions than for systematic and global evaluations of an entire reconstruction effort.

[12] Cf. Willems (1993) "Ich gehöre zu denjenigen, die nicht daran glauben, dass, wir es in der Archäologie mit Tatsachen zu tun haben. Was wir ausgraben, und was wir daraus schliessen, das sind keine Fakten, sondern Interpretationen, geprägt von den heutigen technischen Möglichkeiten und unseren Vorstellungen." – "I belong to those who do not believe that we are dealing with facts in archeology. What we dig up, and what we conclude from it, are not facts but interpretations, circumscribed by current technical possibilities and our imagination." (translation E.K., p. 25).

[13] The first author of this article is an excellent example. Despite extensive phonetic training, early exposure to French, 20 years of daily use and good motivation, he still devoices initial voiced plosives in French, just as he does in his native language of German and his main second language which is English.

[14] All while admitting that some intrinsic limits are inevitable. As in modern languages (e.g., French), the pronunciation of some graphemic forms is inherently ambiguous (e.g., the difference between Latin 'INVENIT "finds" and INV'ENIT "found"). This type of ambiguity can only be resolved by advanced semantic and grammatical processing, which is not generally available in low-cost systems. But our experience with the

synthesis of modern languages suggests that first, such instances are mercifully rare in most languages (Chinese being the obvious exception!), and second, that well-chosen heuristics and mark-up conventions generally resolve such issues with minimal computational overhead. The INVENIT problem could for example be solved by imposing a mark-up containing accent marks preceding the stressed syllable or by a heuristic that attempts to establish the prevailing tense in surrounding verbs.

[15] The authors' reading of the evidence is that native Latin words with word-initial /h/ were indeed aspirated, and that only hypercorrections of loan words were exposed to ridicule (Bianchi, forthcoming). However, this point of view is not universally shared.

[16] Cicero remarks on this phenomenon: "Quin ego ipse, cum scirem ita maiores locutos ut nusquam nisi in uocali aspiratione uterentur, loquebar sic ut *pulcros*, *Cetegos*, *triumpos*, *Cartaginem* dicerem ; aliquando, idque sero, conuicio aurium cum extorta mihi ueritas esset, usum loquendi populo concessi, scientiam mihi reseruaui. *Orciuios* tamen et *Matones*, *Otones*, *Caepiones*, *sepulcra*, *coronas*, *lacrimas* dicimus, quia per aurium iudicium licet." "Myself, knowing that our ancestors placed aspiration nowhere except in vowels, I stuck to the pronunciation *pulcros, Cetegos, triompos, Cartaginem*. At a certain moment quite late, when my ears objected to this, I abandoned my wish to correct, conceded to the people the right to fix usage, and kept my science to myself. It remains that we say *Orciuios, Matones, Otones, Caepiones, sepulcra, corones, lacrima*, as the judgment of the ears suggests." (Cic. *orat.* 160).

[17] It is interesting to note that listening tests using our French speech synthesis have demonstrated a strong sensitivity to the duration of some vowels. In semi-formal tests with various members of our laboratory, we varied the duration of the /i/ vowel in a standard paragraph by between 1% and 10% and asked for *acceptable/unacceptable* judgements. At 2%, listeners reversed judgement for several words, and at 5%, for the majority of words containing the /i/-vowel. At an average duration of some 70 ms for the vowel /i/, 2% average variation works out to less than two milliseconds. We conclude from this and many similar experiments that the reconstruction of precise timing information is not a trivial exercise. Despite inevitable individual and contextual variation, certain sounds in certain phonological contexts may well have characteristic and distinctive durations which are judged according to severe perceptual standards by native speakers. Since ancient languages are no longer spoken natively, the precise perceptual acuity to judge such durational adequacy may be lost forever.

[18] See sound example at www.unil.ch/imm/docs/LAIP/LAIPTTS_F_ex_eng.htm.

[19] See sound examples at www.unil.ch/imm/docs/LAIP/LAIPTTS_L_SpeechMill_dl.htm.

[20] Some promising systematic work in this direction is currently being undertaken in our laboratory (Zellner Keller, work in progress).

[21] Some further inferences may well be made from described physical appearance to voice quality. Santi et al. (2001), for example, presented 15-minute anonymous recordings to listeners and had them score such characteristics as age, sex typicality, weight, size, eye colour, as well as colour and length of hair of the speaker. Preliminary results from six listeners indicate that at least for some speakers, age and sex typicality as well as colour and hair length (!) are judged quite reliably from the vocal quality alone, suggesting that there are some predictable relationships between speakers' socially relevant physical characteristics and their vocal characteristics. Exactly which signal characteristics transmit this information, however, is open to investigation.