Speech-to-text e wiki per l'evoluzione di un servizio di podcasting didattico

Federica Baroni[1], Alberto Betella[2], Marco Lazzari[1]
[1] *Università di Bergamo, Dipartimento di Scienze umane e sociali*
Piazzale Sant'Agostino 2, 24129 Bergamo
{federica.baroni | marco.lazzari}@unibg.it
[2] *Universitat Pompeu Fabra, SPECS Laboratory*
Carrer de Roc Boronat 138, 08018 Barcellona, Spagna
alberto.betella@upf.edu

Si presenta l'evoluzione di una libreria software per la creazione e gestione di servizi di podcasting, per la quale è stata realizzata un'estensione che, al momento del caricamento di un nuovo podcast, ne genera automaticamente una trascrizione testuale caricabile in un sistema wiki che consente la correzione di eventuali errori da parte degli utenti. Ne è risultata una piattaforma integrata con la quale è in corso una sperimentazione in ambienti didattici dalla scuola primaria all'università, per valutare l'impatto sugli apprendimenti degli aspetti interazionali, motivazionali e metacognitivi collegati all'uso di ambienti wiki.

1. Introduzione

Nato nel 2004 e destinato inizialmente all'intrattenimento, il podcasting si è presto diffuso come strumento di didattica a distanza e supporto ad attività di apprendimento orientate al mobile learning [McGarr, 2009]. La ricerca sul podcasting educativo ha preso avvio in modo più consistente nel 2005; in particolare, maggiore attenzione è stata dedicata all'istruzione superiore [Evans, 2008; Hew, 2009], al fine di sviluppare buone pratiche, fornendo nuovi modelli di insegnamento-apprendimento nella realtà scolastica mutata dal digitale [Sharpe et al., 2010], e di definire criteri per implementare servizi di podcasting di qualità [Lazzari e Betella, 2007]. A dieci anni dall'avvio delle sperimentazioni, emerge che il podcasting è considerato uno stimolo all'apprendimento con un significativo potenziale pedagogico [Ralph et al., 2010; Pegrum et al., 2014], in particolare per la possibilità di migliorare lo stile di insegnamento dei docenti e la qualità dei materiali delle lezioni [Brittain et al., 2006; Cebeci e Tekdal, 2006], per sostenere la motivazione degli studenti [Lee et al., 2008; Armstrong et al., 2009] e stimolare l'apprendimento soprattutto quando il podcast è l'esito di un'attività collaborativa nell'ambito di una didattica laboratoriale che coinvolge gli studenti nella produzione dei contenuti [Smith et al., 2005; Lazzari, 2009].

D'altra parte, si cominciano a rimeditare le possibilità applicative del podcasting alla didattica, visti anche talune evidenze contraddittorie [Fernandez et al., 2015] e alla luce delle mutate condizioni delle tecnologie telematiche, legate all'aumentata disponibilità della connettività mobile, la cui diffusione tende a comprimere il ruolo, forse troppo enfatizzato, della fruizione nomadica del podcasting: posto che ormai chiunque può accedere a contenuti multimediali tramite smartphone, qual è l'essenza del podcasting che lo differenzia e lo può far preferire a servizi di streaming o downloading acceduti in mobilità?

Nello stesso decennio è anche aumentato l'interesse per i servizi speech-to-text, che permettono la conversione in testo scritto di un discorso registrato in formato audio o video. L'uso di software stand-alone per questo scopo ha una lunga storia, ma di recente si è avuta una svolta tecnologica importante, con il passaggio dei servizi di trascrizione dal computer locale al cloud. Ciò consente di disporre di più potenza computazionale e di implementare sul server remoto tecniche di machine learning, per migliorare l'accuratezza del riconoscimento vocale grazie alla grande massa di esempi di parlato disponibili. In ambito didattico, che ha esigenze diverse da quello della dettatura di testi, il rinnovato impegno della ricerca ha avuto l'importante ricaduta di proporre servizi che non richiedono addestramento per tararsi sulla voce del parlante. Ciò permette di registrare e convertire in testo lezioni tenute da docenti ospiti, seminari, tavole rotonde, senza obbligare ciascuno speaker a noiose e onerose sedute di addestramento, tollerabili se ammortizzate rispetto a un lungo periodo di uso dello strumento, ma improponibili all'ospite di un evento che arrivi all'ultimo minuto [Bain et al., 2012; Ranchal et al., 2013; Wald e Li, 2012].

Negli stessi anni si sono diffusi servizi di scrittura collaborativa in ambito accademico e più in generale educativo [Parker e Chao, 2007; Brodahl et al., 2011; Tomlinson et al., 2012], in particolare dei sistemi wiki non solo come ambienti in cui reperire facilmente contenuti più o meno corretti [Lazzari, 2014], ma anche come strumenti di apprendimento attivo e collaborativo [Cress e Kimmerle, 2008; Wheeler et al., 2008]. Non mancano voci critiche [Cole, 2009] o quantomeno scettiche [Ebner et al., 2008; Elgort et al., 2008], ma vari studi hanno mostrato le possibili applicazioni dei wiki nella didattica, e in riferimento alla prospettiva del presente lavoro sono documentati esiti incoraggianti rispetto all'uso della scrittura collaborativa via wiki [Biasutti e EL-Deghaidy, 2012; Trentin, 2012; Aydin e Yildiz, 2014] o all'opportunità di incrementare le competenze di scrittura degli studenti universitari quando correzione e valutazione avvengono tra pari, ottenendo livelli di soddisfazione e motivazione maggiori rispetto alle pratiche tradizionali [Xiao e Lucking, 2008].

2. L'integrazione di podcasting, speech-to-text e wiki

Sin dai primi anni della storia del podcasting il nostro gruppo di lavoro vi si è dedicato sia dal punto di vista delle applicazioni didattiche [Lazzari, 2007] e della riflessione critica su di esse [Lazzari e Betella, 2007], sia da quello della

creazione di un ambiente software per la pubblicazione degli episodi di podcasting. Ciò ha portato allo sviluppo, a partire dal 2005, di Podcast Generator (PG), un content management system (CMS) open source dedicato alla pubblicazione di podcast (podcastgen.sourceforge.net). Tramite PG l'utente dispone di tutti gli strumenti per pubblicare episodi audio e video in formato podcast. La libreria software è stata costantemente migliorata e aggiornata secondo le esigenze dell'Università di Bergamo ed è stata installata da migliaia di servizi di podcasting in tutto il mondo, compresi quelli di vari atenei.

Attualmente è in corso di test una estensione che, al momento del caricamento di un nuovo podcast, ne genera automaticamente una trascrizione testuale, che viene caricata in un sistema wiki che consente la correzione di eventuali errori da parte degli utenti. L'obiettivo è proporre trascrizioni utili a chi per abitudine, stile di studio, disabilità sensoriale, specifico bisogno educativo o disturbo attentivo necessita di testi che integrino o sostituiscano l'audio o la partecipazione in presenza, o sostengano e compensino difficoltà apprenditive. Inoltre, con il wiki si intende stimolare virtuosi processi di scrittura collaborativa.

Nel seguito presentiamo i dettagli implementativi della nuova architettura di PG con l'estensione per riconoscimento vocale e scrittura collaborativa. Il piano di lavoro ha previsto che le nuove funzioni venissero testate prima su una versione "ombra" di Pluriversiradio, poi a partire dal secondo semestre dell'anno accademico 2014-15 sulla versione ufficiale e a seguire su una serie di installazioni federate (almeno una scuola primaria, una secondaria di primo grado e una di secondo), con le quali studiare l'impatto educativo delle novità.

Il sistema di riconoscimento del parlato si basa su una serie di script server-side che possono venire lanciati automaticamente (per essere eseguiti dal server) dalla pagina web al caricamento di un contenuto multimediale (tipicamente audio, ma eventualmente anche video). I pacchetti software usati, tutti disponibili sotto licenza open source, sono normalmente inclusi nelle più comuni distribuzioni di Linux e, in alcuni casi, in Mac OS X.

Innanzitutto è necessario convertire il file originale in formato waveform audio (.wav), che non presenta compressione ed è raccomandato qualora si abbia l'esigenza di usare sistemi di riconoscimento vocale in tempo reale (così evitando di consumare potenza di calcolo per la decodifica di formati compressi, come, per esempio, mp3). Nel caso di contenuti video è opportuno estrarre la traccia audio prima di procedere alla conversione. Ciò può essere fatto usando applicazioni quali ffmpeg (www.ffmpeg.org). Poi viene operata una suddivisione in sotto-tracce (audio chunks) della durata di pochi secondi. Molte web API di riconoscimento vocale, infatti, pongono limitazioni rispetto alla durata massima di una singola traccia. Ciò alleggerisce il carico di lavoro dell'API e permette di restituire velocemente il risultato testuale, evitando ritardi dovuti a lunghi tempi di caricamento e offrendo all'utente l'illusione di una trascrizione in tempo reale (a rigore PG non ne avrebbe bisogno, poiché non ha il vincolo della trascrizione

in tempo reale, per la quale invece sono tipicamente pensati i programmi di riconoscimento vocale e speech-to-text).

Spesso in un discorso (si pensi a un docente che fa lezione o a un'intervista) ricorrono numerose pause e silenzi. Pertanto, un utile accorgimento consiste nell'impostare un filtro che riconosca i momenti di silenzio in base a limiti (thresholds) predefiniti e che li escluda dalle tracce inviate all'API, riducendo il peso dei file da caricare e conseguentemente il tempo di attesa dell'utente. Questa tecnica si rivela ancor più utile quando si dispone di una connessione lenta, come accade in caso di bassa copertura di rete mobile. Inoltre, la riduzione della qualità (downsampling) di una traccia audio, specie nel caso di discorsi parlati, dove la qualità non è una variabile critica, può diminuire drasticamente il peso delle tracce. Sia la suddivisione di una traccia madre in sotto tracce, sia il riconoscimento dei silenzi e il downsampling di file audio possono essere operate attraverso l'uso di programmi di sound processing come SoX (sox.sourceforge.net).

Caricata la (sotto)traccia audio, l'API restituisce al client (che può essere un dispositivo mobile, o semplicemente il navigatore web) un file testuale contenente la trascrizione del testo e opzionalmente altri parametri. I formati più usati per lo scambio di questo genere di dati sono XML e JSON, che permettono di etichettare semanticamente la trascrizione, offrendo dunque un'interpretazione più ricca dei risultati. La maggior parte delle API web per il riconoscimento vocale oggi disponibili, infatti, non si limita alla mera restituzione di trascrizioni testuali, bensì arricchisce il contenuto con informazioni quali diverse ipotesi d'interpretazione dell'audio, ciascuna corredata da enunciati (utterance), percentuale di confidenza di trascrizione di ciascuno di essi e, nei casi di API concepite per l'esecuzione di comandi vocali, intenzioni dell'utente.

Per implementare il riconoscimento vocale in Pluriversiradio abbiamo sperimentato diverse piattaforme che avessero come prerequisiti la presenza di documentazione e la possibilità di accesso attraverso client o applicazioni di terze parti. Molti servizi di riconoscimento vocale, infatti, sono fruibili solo tramite servizi specifici come, per esempio, l'API su cui si basa Siri (l'assistente vocale di iPhone e iPad), che è riservata ai prodotti Apple. Abbiamo così individuato tre possibili servizi: Sphinx (cmusphinx.sourceforge.net), un programma open source di speech recognition sviluppato alla Carnegie Mellon University e installabile su un server, la Speech API offerta da Google (code.google.com/apis/console), e Wit.ai (wit.ai), un servizio web di speech-to-text sviluppato da una giovane startup californiana. Sphink è stato escluso dopo i primi test di trascrizione per gli scarsi risultati dovuti alla mancanza di un modello linguistico per l'italiano. L'API di Google ha prodotto ottimi risultati, presentando tuttavia limiti particolarmente restrittivi (almeno nella versione gratuita) di 50 query al giorno con 10 secondi massimi di durata ciascuna. Infine, con Wit.ai abbiamo ottenuto risultati molto soddisfacenti e accurati. Questa API allo stato attuale non presenta limitazioni in termini di query e offre supporto per la lingua italiana. Perciò Wit.ai è stata scelta e integrata nel

sistema di riconoscimento vocale di Pluriversiradio (la bontà della scelta è stata poi confermata dall'acquisizione di Wit.ai da parte di Facebook).

Ultimata l'implementazione della nostra architettura speech-to-text, abbiamo proceduto a una definizione empirica dei parametri ottimali per la fase di post-processing audio che precede la trascrizione. Non esiste attualmente un vero e proprio insieme di parametri / indicatori / metriche di riferimento che possano fungere da linee guida per ottimizzare il file audio originale, dal momento che la qualità della registrazione varia in base all'hardware (per esempio, in funzione della qualità del microfono utilizzato) e alle condizioni ambientali (sala conferenze, aula, registrazioni all'aperto). Per questo motivo stiamo sperimentando diverse combinazioni di threshold per il silenzio, sampling rate e lunghezza delle sotto-tracce, al fine di ottenere file ottimizzati per essere caricati rapidamente, senza pregiudizio della qualità della trascrizione.

Infine, il testo generato dal convertitore viene caricato in uno spazio wiki, in modo che gli utenti lo possano correggere (è scontato che il testo sia impreciso, per motivi che vanno dalla qualità dell'audio alla limitatezza del programma di conversione, passando per l'accento del parlante e la ricchezza del suo lessico), formattare e discutere. Per l'implementazione del wiki siamo ricorsi a DokuWiki (dokuwiki.org), una piattaforma completa e open source che non richiede un database specifico, poiché si basa su semplici file di testo, ciascuno dei quali contiene una pagina. In questo modo le trascrizioni generate dal sistema come file testuali diventano direttamente articoli del wiki, senza ulteriori elaborazioni per l'immagazzinamento in un database.

Le modifiche ai testi saranno possibili, almeno nelle fasi di sperimentazione, soltanto a utenti registrati e noti ai gestori del servizio, per poter monitorare gli effetti della cooperazione sugli apprendimenti e per scongiurare gli atti di vandalismo che spesso affliggono i siti 2.0 (e Wikipedia in particolare).

3. Discussione e sviluppi

La trascrizione dei podcast offre almeno due vantaggi: l'accessibilità dei contenuti e una migliore indicizzazione nei motori di ricerca, che possono usare il corpus stesso della trascrizione di ciascun episodio e non solo parole chiave e meta tags apposti dall'utente (con inevitabile filtro interpretativo). Per quanto riguarda l'accessibilità, in particolare, un sistema automatico di trascrizione come quello descritto favorisce sia persone con disabilità uditiva, sia chi ha scarsa competenza linguistica in una determinata lingua (L2) e trova efficace supporto quando all'ascolto può associare la lettura dei relativi contenuti.

Trascrivere l'audio mette in atto il primo principio dello Universal Design for Learning, che suggerisce di prevedere mezzi alternativi di rappresentazione [Rose e Meyer, 2002]. Dal momento che le soluzioni proposte non si inseriscono come adattamento dedicato e a posteriori, ma sono integrate nel processo iniziale di produzione dei contenuti, si ha garanzia di reale

disponibilità di materiali didattici di qualità e fruibili da tutti. Per le persone sorde le trascrizioni hanno doppia valenza positiva, giacché consentono la fruizione in sé dei contenuti e, in maniera simile a quanto avviene con la sottotitolazione intralinguistica, possono servire a rafforzare la competenza linguistica, spesso lacunosa in persone con deficit uditivo [Maragna et al., 2013]. Questo secondo aspetto, valido anche per studenti in L2, è un esempio dei possibili risvolti didattici dello strumento descritto. È da notare inoltre che, nei colloqui e interviste che hanno preceduto lo sviluppo del nostro sistema, hanno manifestato interesse per le trascrizioni anche gli studenti che abitualmente registrano le lezioni universitarie, per poi sbobinarle a casa: nel loro caso la sbobinatura sarebbe effettuata dal programma di conversione e a loro resterebbe la sistemazione del testo su wiki, per emendarlo da errori e imprecisioni della conversione con una facilitazione della pratica del prendere appunti [Traphagan et al., 2010].

Un'ulteriore ricaduta positiva del sistema nei contesti di istruzione e formazione riguarda la fase di post-produzione: la piattaforma, sfruttando il sistema collaborativo di tipo wiki, consentirà agli utenti di interagire tra loro e di intervenire modificando (migliorando o arricchendo) la trascrizione operata dalla macchina; in questo modo, attraverso una pratica di crowdsourcing, gli studenti sono direttamente coinvolti nella produzione delle lezioni e non semplicemente passivi fruitori, con le auspicabili ripercussioni motivazionali del caso.

Con una ricerca sperimentale, che introduce la piattaforma integrata in realtà scolastiche di diverso ordine e grado, intendiamo verificare se (e come) l'uso del podcasting secondo il modello collaborativo impatta sui processi di insegnamento/apprendimento, con particolare attenzione agli aspetti interazionali, motivazionali e metacognitivi. Adotteremo una metodologia mista (sia di analisi quantitativa sia qualitativa) basata sulla comparazione tra un gruppo sperimentale di studenti che farà un uso attivo del podcasting e un gruppo di controllo che ne farà un uso tradizionale: questionari, interviste e osservazioni sul campo consentiranno di valutare la qualificazione della didattica, l'auto-percezione del processo di apprendimento e il grado di coinvolgimento delle figure interessate, al fine di diffondere buone pratiche per realizzare podcast didattici di qualità, nell'interesse di docenti, tutor e progettisti della formazione.

Bibliografia

Armstrong G. R., Tucker J. M., Massad V.J., Interviewing the experts: Student produced podcast. Journal of Information Technology Education: Innovations in Practice, 8, 2009, 79-89.

Aydin Z., Yildiz S., Using Wikis to promote collaborative EFL writing. Language Learning and Technology, 18, 1, 2014, 160-180.

Bain K., Stevens J., Martin H., Lund-Lucas E., Transcribe your class: Empowering students, instructors, and institutions: Factors affecting implementation and adoption of a

hosted transcription service, in Proc. of the 6th Int. Technology, Education and Development Conf. (INTED2012), Valencia, Spain, IATED, Valencia, Spain, 2012, 1446-1454.

Biasutti M., EL-Deghaidy H., Using Wiki in teacher education: Impact on knowledge management processes and student satisfaction. Computers and Education, 59, 3, 2012, 861-872.

Brittain S., Glowacki P., Van Ittersum J., et al., Podcasting lectures. EDUCAUSE Quarterly, 29, 3, 2006, 24-31.

Brodahl, C., Hadjerrouit, S., Hansen, N.K., Collaborative writing with Web 2.0 technologies: education students' perceptions, Journal of Information Technology Education: Innovations in Practice, 10, 2011, 73-103.

Cebeci Z., Tekdal M., Using podcasts as audio learning objects. Interdisciplinary Journal of Knowledge and Learning Objects, 2, 2006, 47-57.

Cole M., Using Wiki technology to support student engagement: Lessons from the trenches. Computers and Education, 52, 1, 2009, 141-146.

Cress U., Kimmerle J., A systemic and cognitive view on collaborative knowledge building with wikis. Computer-supported Collaborative Learning, 3, 2, 2008, 105-122.

Ebner M, Kickmeie-Rust M., Holzinger A., Utilizing Wiki-Systems in higher education classes: a chance for universal access? Universal Access in the Information Society, 7, 4, 2008, 199-207.

Elgort I., Smith. A.G., Toland J., Is wiki an effective platform for group course work? Australasian Journal of Educational Technology, 24, 2, 2008, 195-210.

Evans C., The effectiveness of m-learning in the form of podcast revision lectures in higher education. Computers and Education, 50, 2, 2008, 491-498.

Fernandez V., Sallan J.M., Simo P., Past, Present, and Future of Podcasting in Higher Education, in Li M. e Zhao Y. (eds) Exploring Learning & Teaching in Higher Education. New Frontiers of Educational Research, Springer, Berlin, 2015, 305-330.

Hew K. F., Use of audio podcast in K-12 and higher education: A review of research topics and methodologies. Education Technology Research and Development, 57, 3, 2009, 333-357.

Lazzari M., Betella A., Towards guidelines on educational podcasting quality: problems arising from a real world experience, in Smith M.J. e Salvendy G. (eds.), Human interfaces, Part II, Berlin, Springer, 2007.

Lazzari M., Creative use of podcasting in higher education and its effect on competitive agency. Computers & Education, 52, 1, 2009.

Lazzari M., Informatica umanistica, McGraw-Hill, Milano, 2014.

Lazzari M., Podcasting in the classroom: involving students in creating podcasted lessons, in Proc. of the Conf. HCI Educators 2007 (HCIEd 2007), Aveiro, Portugal, 2007.

Lee M.J.W., McLoughlin C., Chan A., Talk the talk: Learner-generated podcasts as catalysts for knowledge creation. British Journal of Educational Technology, 39, 3, 2008, 501-521.

Maragna S., Roccaforte M., Tomasuolo E., Una didattica innovativa per l'apprendente sordo, FrancoAngeli, Milano, 2013.

McGarr O., A review of podcasting in higher education: Its influence on the traditional lecture. Australasian Journal of Educational Technology, 25, 3, 2009, 309-321.

Parker, K.R., Chao, J.T., Wiki as a teaching tool, Interdisciplinary Journal of Knowledge and Learning Objects, 3, 2007, 57-73.

Pegrum M., Bartle E., Longnecker N., Can creative podcasting promote deep learning? The use of podcasting for learning content in an undergraduate science unit. British Journal of Educational Technology, 46, 1, 2014, 142-152.

Ralph N., Head N., Lightfoot S., Pol-Casting: The use of podcasting in the teaching and learning of politics and international relations. European Political Science, 9, 2010, 13-24.

Ranchal R., Taber-Doughty T., Guo Y., et al., Using speech recognition for real-time captioning and lecture transcription in the classroom. IEEE Transactions on Learning Technologies, 6, 4, 2013, 299-311.

Rose D., Meyer A., Teaching every student in the digital age, ASCD, Alexandria, VA, 2002.

Sharpe R., Beetham H., De Freitas S., Rethinking learning for a digital age, Routledge, New York, 2010.

Smith K.A., Sheppard S.D., Johnson D.W., et al., Pedagogies of engagement: classroom-based practices. Journal of Engineering Education, 94, 1, 2005, 87-101.

Tomlinson, B., Ross, J., André, P., et al., Massively distributed authorship of academic papers, in CHI '12 Extended Abstracts on Human Factors in Computing Systems, Austin, TX, ACM Press, New York, NY, 2012, 11-20.

Traphagan T., Kusera J. V., Kishi K., Impact of class lecture webcasting on attendance and learning. Educational Technology Research and Development, 58, 1, 2010, 19-37.

Trentin B., Using a wiki to evaluate individual contribution to a collaborative learning project. Journal of Computer Assisted Learning, 25, 1, 2012, 43-55.

Wald M., Li Y., Synote: Important enhancements to learning with recorded lectures, in Proc. of the 12th IEEE International Conference on Advanced Learning Technologies (ICALT 2012), Rome, Italy, IEEE Computer Society, Los Alamitos, CA, 2012, 521-525.

Wheeler S., Yeomans P., Wheeler D., The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. British Journal of Educational Technology, 39, 6, 2008, 987-995.

Xiao Y., Lucking R., The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. Internet and Higher Education, 11, 3, 2008, 186–193.