

Mixed-mode Multilinguality in TTS: The Case of Canadian French

Laura Mayfield Tomokiyo

Cepstral LLC
Pittsburgh, USA

laura@cepstral.com

Carol Jay Sisson

McGill University
Montreal, Canada

carol.sisson@mcgill.ca

Alan W Black

Cepstral LLC
Pittsburgh, USA

awb@cepstral.com

Abstract

The coexistence of English and French in Canada presents a number of interesting problems for text-to-speech (TTS) synthesis. The pronunciation of Canadian French is fairly well documented and can be captured by recording a speaker of the appropriate dialect for the voice database. The desired behavior of the system in speaking the many English words, names, and expressions that can populate French text, however, is not well understood, varying from user to user and from context to context. In this paper we present an analysis of English in Canadian French TTS, examining the intelligibility and preferability of English and French pronunciations. Our results suggest that it is best to consider different *modes* of synthesis, ranging from near-English pronunciation of English terms to near-French, and that different tasks require different approaches to the problem.

1. Introduction

Canada is an officially bilingual country, supporting equal rights for English and French (and in some cases indigenous languages) in government, education, and commerce. The implementation varies by province and municipality, with each province having certain responsibilities to provide bilingual services as well as freedom to adopt measures that further foster the recognition and use of French or English [1]. 18% of Canadians, and 41% of residents of Quebec, report themselves to be fully bilingual in English and French [2].

Official language policy may dictate what language is printed on street signs, official documents, or consumer packaging. How English and French are integrated in natural speech, however, is a different matter. In speech, both bilinguals and monolinguals make instinctive decisions about the pronunciation and modification of terms based on their audience, the situation, and their own experience. Historically, the English terms used in spoken French are not pronounced in a way that is recognizable to a native monolingual English speaker, although they are considered to be English words. It is not clear that they are pronounced strictly according to French letter-to-sound conventions. Rather, it seems to be a hybrid of an English-like letter-to-sound mapping with a further mapping that is part phonetic and in part full phoneme substitution, insertion, and deletion.

In this paper, we explore the role of English terms in French text-to-speech (TTS) unit selection synthesis. In early investigations, we found francophone listeners to be extremely sensitive to the pronunciation of English-origin words, both in terms of comprehension and in terms of the

statement that was being made by generating a pronunciation that was “too English.” We strive to understand how best to render English in different types of text, analyzing both the distribution of English terms and user preference with respect to their pronunciations.

2. Bilingualism in Canada

Among Canadian provinces, only New Brunswick is officially bilingual. The province of Quebec is officially French, and the city of Ottawa is officially bilingual. All provinces have some French-speaking community, although English speakers are in the minority only in Quebec. Both the mandate and custom of multilinguality, however, vary by province, municipality, and community, as well as situation.

Streets in Ottawa, for example, have official French and English names. In Montreal, some boroughs do have official French and English names for streets, with the signs displaying both designations (e.g. RUE SHERBROOKE ST). Cross-lingual equivalents of geographical names aren’t always clear, however. Would an English speaker say “Avenue de l’Eglise,” or would they change it into “de l’Eglise Avenue,” “Eglise Avenue,” or even “Church Avenue?” And would they make the same choices if they were reading a street sign or instructing someone to “turn left on Avenue de l’Eglise?” Much highway signage is bilingual.

Education, while primarily French in Quebec and English elsewhere, is provided for minority communities where numbers warrant. French-language schooling is available for groups of as few as 15 students in parts of Canada. Children of Canadian citizens desiring English education in Quebec must meet eligibility requirements; immigrants attending public schools in Quebec must attend French-speaking schools.

In the media, publications and television channels are identified as French or English; 4 French and 3 English broadcast channels are available in Montreal, for example, including French and English versions of the Canadian Broadcasting Company (CBC). English-language newspapers have style guidelines for translating French terms, but these guidelines can vary from organization to organization. English/French alternations are sometimes seen in text; for example, as we might see “he/she” in an American text, we might find “Canadians/Canadien(ne)s” in a Canadian text. Quotations are often included in the source language, although with translations. In cases like these where two languages are provided, one task of TTS front-end processing is to decide whether the expressions should be presented in their entirety or monolingually when bilingual content is redundant. Many websites, and all

official government sites, have either parallel text presented on the same page or parallel portals. Radio stations may be flexible with language use, and of course many of the songs and musical groups announced on francophone popular music stations have English names.

Attitudes toward multilinguality and relationships between francophone and anglophone communities vary across Canada, with tensions being highest in Quebec and particularly in Montreal. Many speakers identify strongly with their language environment. They may have negative associations with other language communities, and may not respond well to incorrect or uninformed assumptions about language use.

Prominent dialects of French spoken in Canada include Quebecois, Franco-Ontarian, and Acadian. Overall, 59% of Canadians are anglophone and 23% francophone, with 18% having another mother tongue. Significant among this group are speakers of Chinese, Italian, Arabic, and indigenous languages [2].

3. Multilinguality in TTS

Most TTS voices are monolingual; that is, they expect input text in a given language and make no special attempt to either identify or provide special handling for regions of text in a different language. Even if the lexicon includes terms of foreign origin, in most cases the voice database is monolingual and pronunciations involving sounds outside the normal phonetic inventory of the voice language cannot be generated.

Multilingual demands on TTS can be divided into several categories.

3.1. Lexicalized foreign-origin terms or names

As long as language communities have interacted, there has been migration of words between languages. As words come into common use, their pronunciations and usage become fixed; these are called *lexicalized* terms. Most common lexicalized terms might be expected to appear in their base form in a good pronunciation dictionary. In languages more highly inflected than English, we also see a variety of related forms, where the base word retains its original spelling and approximate pronunciation but the inflected parts are in the native language and use a native letter-to-sound relationship. For example, in German, the word “update” has the same meaning as it does in English, and approximately the same pronunciation (/ʌpdeɪt/). We also see inflected forms such as *geupdatet* (past tense) and *updaten* (infinitive), which may not be in the lexicon. The challenge for the TTS system in this situation is to identify not only the word but the part of the word that is of foreign origin, and apply different letter-to-sound rules for the native and non-native portions [3].

3.2. Novel foreign-origin terms or names

When we read about current events, we frequently encounter foreign names, both of people and geographical locations, that are not yet in common use and that there may not be agreement on how to pronounce. *Kirkut*, *Chavez*, and *Banda Aceh* are examples of the foreign names that at the time of this study were frequently in the news. Most educated English speakers recognize them easily in text, but may not have had occasion to pronounce them. To then synthesize these names in a way that is consistently understandable is a significant challenge

for TTS.

3.3. Interleaved regions of text

Code switching, or switching between languages or language varieties [4], is not uncommon in communication between members of a multilingual community. When a speaker is fluent in more than one language, they may find that a word or phrase from one language expresses what they want to say better than one that is available in the language being spoken. Instances of code-switching can range from a single word to a complete shift in the primary language of the interaction. Code-switching is primarily a characteristic of spoken language, and has not traditionally been common in text. As e-mail, blogs, and other forms of online text represent a form of written language that is closer to spoken language, however, we see writers interleaving expressions from other languages in situations where the audience is expected to understand them. Identifying these regions is, of course, one problem for TTS. Another substantial challenge, however, is deciding how closely to approximate the language being switched to. In the case of a complete switch of language, it may be best to closely approximate the pronunciation of the new language, to the extent possible within the phonetic space available in the voice. Switching abruptly for short regions, however, may make the overall utterance difficult to understand. Furthermore, as code-switching in text can be used in communities that would not ordinarily code-switch in conversation, it may not be the case that the second language is as well understood in synthesis as it might ordinarily be between fully multilingual conversation partners.

3.4. Parallel regions of text

Displaying of the same text content in more than one language is common in Canada. It is particularly visible in consumer packaging, which must by law present information in both English and French, and on the web, where juxtaposition of English and French text is often seen. Although the language identification problem is present, these texts are often relatively long and provide good opportunity to determine the source language automatically. As the texts are not typically intended for the same audience, it may be acceptable to use completely different voices, one native to each language displayed.

3.5. Accented speech

The four cases described above assume that the text to be synthesized is multilingual to some degree, and that synthesizing it requires switching between linguistic models or voices. A fifth, somewhat separate, challenge is to synthesize monolingual text with a voice that is not native to the target language. In this case, some degree of accentedness is desirable, perhaps to establish a character or identify with a listener group. Methods for generating accented voices could include combining models in an HMM-based system such as [5] and using a hybrid linguistic model [6].

4. Experimental environment

Our objective in this study is to answer the following questions:

1. Where do English terms occur in French text?
2. How do users prefer to hear them pronounced?
3. How can synthesis intelligibility be maximized?
4. Are the answers to these questions domain-dependent?

This section describes the experimental environment, including data and voice resources. The experiments themselves are presented in Section 5.

4.1. Text data

To answer the first question, we examined three different types of text that can reasonably be expected to contain English terms: parliamentary transcripts, online text, and navigation.

4.1.1. Identification of English words

Classification of terms was done on a word-by-word basis. Words were determined to be English if they appeared in either the English TTS lexicon or an English wordlist of approximately 150,000 words *and* did not appear in either the French TTS lexicon or a French wordlist, also of approximately 150,000 words. The wordlists were derived from large corpora of out-of-copyright literature that are expected to be completely monolingual outside of true loanwords; most words that have come into common use since the publication of these literary works are expected to appear in the TTS lexicons, which reflect contemporary usage. This system is not perfect, as it misses many English proper names and also picks up homographs that by coincidence do not appear in the French lexicon. The latter, of course, is also a problem for other methods, such as character n-gram models.

4.1.2. Parliamentary transcripts

Transcripts of parliamentary proceedings were selected because they contain a large number of proper names. Both the Canadian parliament (Hansard) [7] and European parliament (Europarl) [8] data are available in translated parallel corpus form. We elected to use the European parliament data for ease of processing. The Canadian parliament data could reasonably be expected to contain more names that are familiar to Canadians, but the corpora are large enough that extracting appropriate stimuli from the European data was not a problem. The size of the French Europarl corpus used was 17.6 million words.

4.1.3. Online text

A major target of this study was online texts such as e-mail. While a corpus of Canadian French personal e-mail is not publicly available, we were able to build a corpus of online blog, Usenet, and message board text. Posts were determined to be francophone Canadian if they were classified as French using a combination of document- and phrase- level character 4-gram models and contained certain Canadian keywords. One of the difficulties in using language classification on the online data was inconsistent use of accents. For example, the words *meme* (même), *tres* (très), and *deja* (déjà) are all common enough in English to appear in the English lexicon, and did not appear in the French lexicon or wordlist without accents.

The size of the online corpus was 8.2 million words.

Corpus	# of tokens	% of English tokens
Europarl	17.6 million	0.54
Online	8.2 million	1.48

Table 1: Frequency of English words in French text

4.1.4. Navigation

Navigation is a major application for TTS, and a Canadian French navigation system must deal regularly with English geographical names. There is no common corpus for navigation data, but it is easy to generate directions using online mapping sites. For this study, 80 navigation sentences were generated using the Canadian French version of Yahoo Maps to travel across Montreal, Ottawa, and the Maritime provinces. Canadian French localization was somewhat inconsistent at the time of writing, and some manual modifications were necessary (for example, *Continuez sur McKenzie King BR* was changed to *Continuez sur PONT McKenzie King*). Navigation sentences were used only for testing, not corpus analysis.

4.2. Synthesis

Cepstral’s SwiftTMTTS engine was used for all experiments. Swift is a unit selection synthesizer.

The voice used was Cepstral’s Isabelle voice. Isabelle is a Canadian French voice. The model speaker is francophone and from Ottawa. During voice database collection, she was instructed to pronounce any English terms as she naturally would in conversation with other francophones. The promptlist used for collection used a combination of Europarl and Hansard data, as well as general-domain French sentences. Sentences containing English words were not explicitly excluded from promptlist selection, although flagging of foreign-language text is customarily a part of promptlist filtering for other voices.

5. Experiments

5.1. Distribution of English words

The frequency of English word tokens, given in Table 1 was found to be almost three times as high in online data than in the parliamentary transcripts.

The measured numbers for the online text show an occurrence rate of 3 in 200; with an average sentence length of 18 words, this can be thought of one sentence in ten containing an English word. Although the frequency is much lower for the parliamentary data than the online text, there are still nearly ten times more English terms in French Europarl than there are French terms in English Europarl (.54% vs. .06%). We did not formally analyze the distributions in English text, but visual inspection suggests that one factor may be the tendency to leave multi-word expressions, such as organization names, in English in the French data. “Children’s Miracle Network” and “gender mainstreaming” are examples.

5.2. Listening tests

The web-based listening tests described in this paper sought to answer two questions: given a choice, would speakers prefer a sentence with French-like pronunciation of an English word

US	CF	example	English mode	French mode
ɔɪ	o j	playboy	/'pleboj/	/ple'boj/
ɪ	i	Clinton	/'klɪntən/	/klin'ton/
eɪ	e	rains	/'reɪnz/	/'rɛ̃/
æ	a	panthers	/'pænθəz/	/pan'tɛr/
ʌ	ə	from	/'frɒm/	/'frɔ̃m/
aʊ	a w	cowboys	/'kawbojz/	/ka'boj/
aɪ	a j	environment	/ɪn'vejənmənt	/ɔ̃virɔ̃'mɔ̃/
ð	z	southern	/'sʌzðən/	/su'tɛrn/
h	ʔ	Hastings	/'ʔɛstɪŋz/	/'ʔas'tɪŋz/

Table 2: Phone mappings used for generating English-mode pronunciations. The US column shows the standard English phone used in the word and the CF column shows what it was mapped to for English mode. For each example, the actual English- and French-mode pronunciations used in synthesis are shown. Note that the French-mode pronunciations are generated directly from the French lexicon and letter-to-sound model, with no knowledge of English morphology.

or English-like pronunciation, and b) how intelligible are the French/English renderings?

5.2.1. Determination of pronunciations

Although there is expected to be a continuum of pronunciations, ranging from entirely English-like to heavily French-influenced, found in natural speech, this study examines the two extremes. The reason for this simplification is that our Canadian French speech database contains examples of French words and English words pronounced as English within a French sentence as spoken by a single speaker. The actual range of realizations possible in straight unit-selection TTS is dependent on the model speaker's own pronunciations. Our speaker tended toward a relatively English-like pronunciation of common proper names, and so generating a smooth English-like pronunciation was possible in many cases.

Specifically, we compare pronunciations of words as generated by the native French letter-to-sound (LTS) rules with those generated using English LTS rules with a phoneme-level mapping for phones that did not occur in the voice database. We term these English and French *modes*. Mappings used for the English mode are shown in Table 2. The English sounds [ɕ], [θ], [ð], [ɪ], and [v] were used consistently enough by the model speaker and labeled consistently enough in the unit database that they did not need a mapping.

5.2.2. Test subjects

Ten native speakers of Canadian French participated in listening experiments. The age range was 20-29 years. All were students at McGill University and reported that their English comprehension was good to excellent. Subjects were recruited via a McGill job posting resource. For all subjects, the language of primary and secondary education was French and the language of postsecondary education (university) was English. A few of the responses in feedback forms suggested that the writer had a background in linguistics.

All subjects but one were located in Montreal. All completed the tests independently via the internet.

5.2.3. Preference Tests

In a preference test, the listener is presented with two synthesis alternatives and asked to choose which they prefer. This allows us to directly compare French and English modes. Our web-based preference tests display the text being synthesized; we recognize that arguments can be made both for and against displaying the text. Subjects are permitted to listen to the alternatives as many times as they wish. When they have decided, they click on the alternative that they prefer.

Preference tests were completed for all three domains: parliamentary transcripts, online data, and navigation. A fourth preference test on promptlist data was also included. Tests are described in the following paragraphs. Results for all are summarized in Table 3.

Parliamentary Transcripts

40 sentences containing English words were selected from the parliamentary corpus for listening tests. Sentences were automatically filtered for length and ranged from 5 to 15 words. Sentences containing proper names of non-English origin were excluded. For example:

Je considère également cela comme un élément constructif important pour le *gender mainstreaming*.

Back to basics doit être la règle de conduite.

Il ne s'appelle donc pas du tout EEB: *Everything but bananas*.

Navigation

34 navigation sentences were generated using the Canadian French version of Yahoo Maps, as described in Section 4.1.4. Montreal street names were not used because those streets that do have English names (e.g. Sherbrooke) typically have fixed pronunciations within the francophone community and there is not really a question of how they should be pronounced. What we wished to simulate with this test is the experience a driver would have when using a car navigation system to find his way in an unfamiliar city. Street names from Ottawa and Nova Scotia were selected. Names that had the same pronunciations in English and French modes were excluded. For example:

Tournez à gauche sur rue Elgin.

Tournez à droite sur rue Hastings.

Tournez à droite sur avenue Nantucket.

Online text

40 sentences from the online text corpus were selected for the preference test. As with the parliamentary data, sentence length was limited to between 5-15 words. No other exclusions were made, however, as all sentences were naturally-occurring Canadian French text. For example:

Non, je ne parle pas de l'élection de *George Bush*.

The Panthers, beaucoup critiquerons, mais moi j'adore.

Cowboys and Aliens : la réplique belge du stone rock.

Promptlist sentences

To provide a reference for transcription tests to be described in Section 5.2.4, preference tests were performed on sentences including English terms that appear in the voice database. The reasoning for looking at sentences in the database is described in much greater detail in Section 5.2.4. The key point for the preference test, however, is that resynthesizing promptlist sentences greatly reduces the number of synthesis errors and artifacts and allows us to generate speech that is very close to what the model speaker said. 20 promptlist sentences were selected for preference testing. For example:

Car *Steve Jobs* est le roi des présentations devant de larges foules.

Bruce Springsteen, Eminem et la jeune *Norah Jones*.

Dix enfants venant de toutes les régions du Canada ont été désignés champions du *Children's Miracle Network*.

Results for all listening tests are summarized in Section 6.

5.2.4. Transcription tests

In preference tests, listeners are able to express which of two renderings is more pleasing to them, but this does necessarily help us understand how intelligible the synthesis is. In a *transcription test*, the listener is asked to transcribe what they hear, and the results are compared to a reference transcription. We can see how many of the words were understood as well as which ones.

In this study, we wished to test transcription accuracy of both isolated words and words in context. We also wanted separate intelligibility problems due to inappropriate mode from intelligibility problems due to bad synthesis. We chose to address this latter concern by using English terms that appeared in the voice database. We also stepped back from using single isolated words because these are particularly difficult to synthesize well. We instead presented listeners with two types of stimuli: sentences taken directly from the model speaker's promptlist, and the same English words placed in the carrier sentence "Maintenant je dis *word* encore."

It is important to realize that resynthesizing sentences that occurred in the promptlist is not the same as pulling recordings directly out of the database. Although the optimal sequence of synthesis units may well include many or most of the units from the source sentence, this is not always the case, and there are often synthesis artifacts. Furthermore, in this particular data, some of the phones in the English words were excluded from the unit database because they could not be labelled accurately as either French or English; this means that the optimal unit sequence has further opportunity to depart from the source unit sequence.

As the listener transcriptions are scored automatically (the *sclite* [9] tool from NIST was used for scoring), the expected ability of the speaker to understand and accurately spell the words they hear has to be considered. Of the 149 sentences in the database that contained English terms, only 20 were determined to be appropriate for this task. The majority of the rejected sentences contained out-of-date proper names. For example, the name *Newt Gingrich* appeared a number of times in the part of the promptlist based on news from the mid-1990s. We did not feel that a Canadian college

Domain	French mode	English mode
Navigation	55%	45%
Parliamentary	40%	60%
Online	33%	67%
Promptlist	25%	75%

Table 3: Percent of sentences for which French- or English-mode pronunciations were judged to be preferable for the four different domains.

student could be expected to remember, much less accurately spell, the unusual name of a US politician who was in the news when they were ten years old.

The total number of natural-sentence stimuli was 20 and the total number of carrier-sentence stimuli was 30. Results of the transcription tests are given in Table 4.

6. Results

6.1. Preference tests

Results for all preference tests are summarized in Table 3.

It is immediately evident from the test results that listeners prefer different modes for different domains. For the navigation domain, 55% of the sentences were judged to sound better in French mode, followed by 40% for the parliamentary domain, 33% for the online domain, and only 25% for the promptlist sentences.

Excluding for a moment the promptlist sentences, let us consider why listeners might show such responses to the other three domains.

The English words in the navigation sentences were the most likely to be unfamiliar to the listeners, as they involved geographical names in places that they may never have visited. Listeners may be keying strongly off their native French expectations for pronunciations. There is also less semantic content in the surrounding sentence that could aid in interpretation of an unfamiliar name. The presence of the street designation (*street, road*, etc.) may also cause the listener to feel that the designation and name form a single semantic unit that should not be broken up by switching modes.

For the parliamentary data, listeners do seem to prefer the English mode. English terms in this data are most likely names and phrases that they have heard in the news, possibly in both English and French contexts. The relative weakness of the preference may reflect an ambivalence on the part of the listeners as to how the term should be pronounced.

Listeners make a much stronger statement about the preferred mode for the online data. It is our conjecture that the English terms that appear in the online data are more likely to be used by the listeners themselves, both in speech and in text, than those found in the parliamentary data. There is therefore a clearer opinion as to what the best pronunciation is, and although pure English mode may not be the answer, a number of subjects commented that the French-mode pronunciations were "much worse than an ordinary francophone's pronunciation." This seems to be an area where a more intermediate mode might be even more successful.

Let us now return to the promptlist test. The content of the sentences being synthesized is very similar to the parliamentary data, but the preferences shown for English modes are much stronger. As the only real difference is the qual-

Context	French mode	English mode
In context	74.8%	73.4%
In carrier phrase	40.3%	19.3%

Table 4: Transcription accuracy for French- and English-mode synthesis of English terms in and out of context.

ity of the synthesis, we conclude that when the pronunciation is good, listeners favor an English-like rendering of familiar words, but when pronunciation is bad they start to be more comfortable with a French interpretation, even if it is not an accurate representation of how a francophone speaker would pronounce the word.

6.2. Transcription tests

Results for transcription tests are summarized in Table 4. Accuracies are quoted in terms of word accuracy as used in speech recognition. The NIST sclite [9] tool was used for scoring transcription tests. For scoring words in the carrier sentence, only accuracy of the word in question was scored; once listeners figured out what the carrier sentence was, they almost always transcribed it correctly, skewing the scores.

From these results, we see that transcription accuracy of words in their natural sentence context is almost the same for English and French modes, but accuracy for words in a carrier sentence is much worse in English mode. One contributing factor is synthesis quality - the natural sentence contexts appeared in the voice database while the carrier phrase did not. Even considering this, however, it seems clear that listeners are better able to recognize words synthesized in French mode when they cannot rely on contextual clues.

7. Discussion

From the results shown in Section 6, we have drawn several conclusions.

- Listeners prefer different pronunciation modes for different tasks.
- Listeners’ preference for an English-like pronunciation of English terms diminishes with their familiarity with the word.
- Although listeners may *prefer* an English-like pronunciation, they are better able to *recognize* a French-like pronunciation when synthesis is poor or they have limited context.

These conclusions would suggest that it is necessary to provide a range of modes, including not only the extremes but also something in between, for optimal intelligibility and listener acceptance.

Although participants in listening tests were able to assign preference and transcribe what they heard fairly well, they almost universally were dissatisfied with the synthesis of English terms. They were not given any background on the difficulty of the problem or the motivations for synthesizing text in the way that we did, and a number asked why we didn’t just use an English voice. It is clear that the problem is far from solved, and it also seems that it will not be immediately obvious to francophone users of TTS why words that seem commonplace to them are rendered so poorly.

The listeners in this study were comfortable with the idea of completely English pronunciation of English terms and in fact recommended it in their free-form feedback. It is not clear how representative these listeners are of the larger francophone community, however. Preliminary tests and informal conversations indicated that there is a danger of both offending listeners with a pronunciation that is “worse than an ordinary francophone speaker would say it” and alienating listeners with a pronunciation that is “too English.” There seems to be a clear need to explore intermediate layers of pronunciation in future work.

8. References

- [1] Government of Canada, Department of Canadian Heritage, “Official Languages: Myths and Realities,” Hull, Québec, 1998, <http://www.pch.gc.ca/progs/lo-ol/pubs/mythes/english/index.html>.
- [2] Government of Canada, “Profile of languages in Canada: English, French, and many others,” Ottawa, 2002, Statistics Canada 2001 Census Analysis Series, Catalogue no. 96F0030XIE2001005.
- [3] B. Pfister and H. Romsdorfer, “Mixed-Lingual Text Analysis for Polyglot TTS Synthesis,” in *Proc. Eurospeech*, 2003.
- [4] R. Lippi-Green, *English with an Accent: Language, Ideology, and Discrimination in the United States*. London: Routledge, 1997.
- [5] J. Latorre, K. Iwano, and S. Furui, “Cross-language Synthesis with a Polyglot Synthesizer,” in *Proc. Interspeech*, 2005.
- [6] L. M. Tomokiyo, A. W. Black, and K. A. Lenzo, “Foreign Accents in Synthetic Speech: Development and Evaluation,” in *Proc. Interspeech*, 2005.
- [7] LDC, 1997, <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC95T20>.
- [8] P. Koehn, “Europarl: A multilingual corpus for evaluation of machine translation,” 2000, <http://www.isi.edu/%7Ekoehn/publications/europarl.ps>.
- [9] NIST, 2000, <http://www.nist.gov/speech/tools/>.