Example-based Rough Translation for Speech-to-Speech Translation

Mitsuo Shimohata **Eiichiro Sumita**
ATR Spoken Language Translation
Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun
Kyoto 619-0288 Japan
mitsuo.shimohata@atr.co.jp
eiichiro.sumita@atr.co.jp

Yuji Matsumoto
Nara Institute of
Science and Technology
8916-5 Takayama, Ikoma
Nara 630-0101
matsu@is.aist-nara.ac.jp

Abstract

Example-based machine translation (EBMT) is a promising translation method for speech-to-speech translation (S2ST) because of its robustness. However, it has two problems in that the performance degrades when input sentences are long and when the style of the input sentences and that of the example corpus are different. This paper proposes example-based rough translation to overcome these two problems. The rough translation method relies on "meaning-equivalent sentences," which share the main meaning with an input sentence despite missing some unimportant information. This method facilitates retrieval of meaning-equivalent sentences for long input sentences. The retrieval of meaning-equivalent sentences is based on content words, modality, and tense. This method also provides robustness against the style differences between the input sentence and the example corpus.

1 Introduction

Speech-to-speech translation (S2ST) technologies consist of speech recognition, machine translation (MT), and speech synthesis (Waibel, 1996; Wahlster, 2000; Yamamoto, 2000). The MT part receives speech texts recognized by a speech recognizer. The nature of speech causes difficulty in translation since the styles of speech are different from those of written text and are sometimes ungrammatical (Lazzari, 2002). Therefore, rule-based MT cannot translate speech accurately compared with its performance for written-style text.

Example-based MT (EBMT) is one of the corpus-based machine translation methods. It retrieves examples similar to the input sentence and modifies their translations to obtain the output (Nagao, 1981). EBMT is a promising method for S2ST in that it performs robust translation of ungrammatical sentences and requires far less manual work than rule-based MT.

However, there are two problems in applying EBMT to S2ST. One is that the translation accuracy drastically drops as input sentences become long. This is because as the length of a sentence becomes long, the number of retrieved similar sentences greatly decreases. The other problem arises due to the differences in style between the input sentences and the example corpus. It is difficult to acquire a large volume of natural speech data since it requires much time and cost. Therefore, we cannot avoid using a corpus with pseudo speech-style text, which has a little different style from that of natural speech. This style difference makes retrieval of similar sentences difficult and degrades the performance of EBMT.

This paper proposes example-based rough translation to overcome the above two problems of EBMT. Example-based rough translation is characterized by two points: (1) it allows missing unimportant information, and (2) it retrieves similar sentences based on content words and information of modality and tense. Tolerance of missing unimpor-

tant information brings robustness to the translation of long input sentences since this retrieval method substitutes similar short sentences for similar long sentences if there is no similar long sentence. Retrieval based on content word, modality, and tense brings robustness to the style difference between the input sentences and the corpus. The style differences often appear in function words, and this retrieval strategy disregards almost all the information of function words except for the modality and tense information.

We describe the difficulties of applying EBMT to S2ST in Section 3. Then, we describe our purpose and retrieval method for meaning-equivalent sentences in Section 4 and a modification of the translation of meaning-equivalent sentences in Section 5. We report an experiment comparing our method with two other methods in Section 6. The experiment demonstrates the robustness of our method to the length of the input sentence and the style differences between the input sentences and the example corpus.

2 Related Work

The rough translation proposed in this paper is a type of EBMT (Sumita, 2001; Carl, 1999; Brown, 2000). The basic idea of EBMT is that sentences similar to the input sentences are retrieved from an example corpus and their translations become the basis of outputs. Here, let us consider the difference between our method and other EBMT methods by dividing similarity into a content-word part and a function-word part. In the content-word part, our method and other EBMT methods are almost the same. Content words are important information in a similarity measure process, and thesauri are utilized to extend lexical coverage. In the function-word part, our method is characterized by disregarding function words, while other EBMT methods still rely on them for the similarity measure. In our method, the lack of function word information is compensated by the semantically narrow variety in S2ST domains and the use of information on modality and tense. Consequently, our method gains robustness with regard to length and the style differences between the input sentence and the example corpus.

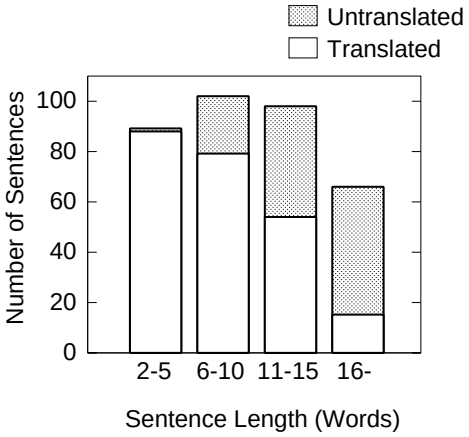

Figure 1: Distribution of Untranslated Input Sentences by Length

3 Difficulties of Applying EBMT to S2ST

3.1 Translation Degradation by Input Length

A major problem with machine translation, regardless of the translation method, is that performance drops rapidly as input sentences become longer. For EBMT, the longer that input sentences become, the fewer similar example sentences exist in the example corpus. Figure 1 shows translation difficulty in long sentences in EBMT (Sumita, 2001). The EBMT system is given 591 test sentences and returns translation results as translated/untranslated. Untranslated means that no similar example sentence exists for the input sentence. Although this EBMT system was equipped with a large example corpus (about 170K sentences), it often failed to translate long input sentences.

3.2 Style Differences between Concise and Conversational

The performance of example-based S2ST greatly depends on the example corpus. It is advantageous for an example corpus to have a large volume and the same style as the input sentences. A corpus of texts dictated from conversational speech is favorable for S2ST. Unfortunately, it is very difficult to prepare such an example corpus since this task requires laborious work, such as speech recording and speech transcription.

Therefore, we cannot avoid using a pseudo speech-style corpus, such as phrasebooks, to prepare

	Language	
	English	Japanese
Concise	5.4	6.2
Conversational	7.9	8.9

Table 1: Number of Words by Sentences

		Language Model	
		Concise	Conversational
Test	Concise	16.4	58.3
	Conversational	72.3	16.3

Table 2: Cross Perplexity

a sufficiently large volume of examples. These texts do not come from real speech but are directly written by imaging speech. They rarely contain unnecessary words. We call the style used in such a corpus "concise" and the style seen in conversational speech "conversational."

Table 1 shows the average numbers of words in concise (Takezawa et al., 2002) and conversational corpora (Takezawa, 1999). Sentences in conversational style are about 2.5 words longer than those in concise style in both English and Japanese. This is because conversational style sentences contain unnecessary words or subordinate clauses, which have the effects of assisting the listener's comprehension and avoiding the possibility of giving the listener a curt impression.

Table 2 shows cross perplexity between concise and conversational corpora (Takezawa et al., 2002). Perplexity is used as a metric for how well a language model derived from a training set matches a test set (Jurafsky and Martin, 2000). Cross perplexities between concise and conversational corpora are much higher than the self-perplexity of either of the two styles. This result also illustrates the great difference between the two styles.

4 Retrieving Meaning-equivalent Sentences for Rough Translation

In order to overcome the problems described in Section 3, we introduce an example-based rough translation strategy. Example-based rough translation has two key features: first, it uses a "meaning-equivalent sentence" which has a looser definition than the conventional "similar sentence" and second, it retrieves meaning-equivalent sentences based on content words and information on modality and tense.

4.1 Meaning-equivalent Sentences

Meaning-equivalent sentences to an input sentence are defined as follows.

> A sentence that shares the main meaning with the input sentence despite missing some unimportant information. It does not contain information additional to that in the input sentence.

They bring robustness to the translation of long input sentences since sentences far shorter than input sentences can be retrieved as meaning-equivalent sentences. We assume that meaning-equivalent sentences (and their translations) are useful enough for S2ST, since unimportant information rarely disturbs the progress of dialogs and can be recovered in the following dialog if needed.

Important information is subjectively recognized mainly due to one of two reasons: (1) It can be guessed from the general situation, or (2) It does not add significant information to the main meaning.

Figure 2 shows examples of unimportant/important information. The information to be examined is written in bold. The information "*of me*" in (1) and "*around here*" in (3) can be guessed from the general situation, while the information "*of this painting*" in (2) and "*Chinese*" in (4) would not be guessed since it denotes a special object. The subordinate sentences in (5) and (6) are regarded as unimportant since they have small significance and are omittable.

4.2 Basic Idea of Retrieval of Meaning-equivalent Sentences

The retrieval of meaning-equivalent sentences depends on content words and basically does not depend on function words. Independence from function words brings robustness to the difference in styles.

However, function words include important information for sentence meaning: the case relation of content words, modality, and tense. Lack of case relation information is compensated by the nature

	Input Sentence	Unimportant?
1	Would you take a picture **of me**?	Yes
2	Would you take a picture **of this painting**?	No
3	Could you tell me a Chinese restaurant **around here**?	Yes
4	Could you tell me a **Chinese** restaurant around here?	No
5	My baggage was stolen from my room **while I was out**.	Yes
6	Please change my room **because the room next door is noisy**.	Yes

Figure 2: Examples of Unimportant Information

of the restricted domain. A restricted domain, as a domain of S2ST, has a relatively small lexicon and meaning variety. Therefore, if content words included in an input sentence are given, their relation is almost always determined in the domain. Modality and tense information is extracted from function words and utilized in classifying the meaning of a sentence (described in Section 4.3.2).

This retrieval method is similar to information retrieval in that content words are used as clues for retrieval (Frakes and Baeza-Yates, 1992). However, our task has two difficulties: (1) Retrieval is carried out not by documents but by single sentences. This reduces the effectiveness of word frequencies. (2) The differences in modality and tense in sentences have to be considered since they play an important role in determining a sentence's communicative meaning.

4.3 Features for Retrieval

4.3.1 Content Words

Words categorized as either noun[1], adjective, adverb, or verb are recognized as content words. Interrogatives are also included. Words such as particles, auxiliary verbs, conjunctions, and interjections are recognized as function words.

We utilize a thesaurus to expand the coverage of the example corpus. We call the relation of two words that are the same "identical" and words that are synonymous in the given thesaurus "synonymous."

[1]Number and pronoun are included.

Modality	Clues
Request	*tekudasai* (auxiliary verb) *teitadakeru* (auxiliary verb)
Desire	*shi-tai* (expression) *te-hoshii* (expression) *negau* (verb)
Question	*ka* (final particle) *ne* (final particle)
Negation	*nai* (auxiliary verb or adjective) *masen* (auxiliary verb)

Tense	Clues
Past	*ta* (auxiliary verb)

Table 3: Clues for Discriminating Modalities in Japanese

4.3.2 Modality and Tense

The meaning of a sentence is discriminated by its modality and tense, since these factors obviously determine meaning. We defined two modality groups and one tense group by examining our corpus. The modality groups are ("request", "desire", "question", "confirmation", "others") and ("negation","others"). The tense group is ("past", "others"). These modalities and tenses are distinguished by surface clues, mainly by particles and auxiliary verbs. These distinguishing rules were manually developed in several weeks. Table 3 shows some of the clues used for discriminating modalities in Japanese. Sentences having no clues are classified as "others". Figure 3 shows sample sentences and their modality

Sentence[2]	Modality & Tense[3]
hoteru *o* yoyaku *shi* tekudasai (Will you reserve this hotel?)	request
hoteru *o* yoyaku *shi tai* (I want to reserve this hotel.)	desire
hoteru *o* yoyaku *shi mashi ta ka*? (Did you reserve this hotel?)	question past
hoteru *o* yoyaku *shi tei masen* (I do not reserve this hotel.)	negation

Figure 3: Sentences and Their Modality and Tense

and tense. Clues are underlined.

A sentence that satisfies the conditions below is recognized as a meaning-equivalent sentence.

4.4 Retrieval and Ranking

1. It has the same modality and tense as the input sentence.

2. All content words are included (identical or synonymous) in the input sentence. This means that the set of content words of a meaning-equivalent sentence is a subset of the input sentence.

3. At least one content word is included (identical) in the input sentence.

If more than one sentence is retrieved, we must rank them to select the most similar one. We introduce "focus area" in the ranking process to select sentences that are meaning-equivalent to the main sentence in complex sentences. We set the focus area as the last N words from the word list of an input sentence. N denotes the number of content words in meaning-equivalent sentences. This is because main sentences in complex sentences tend to be placed at the end in Japanese.

The retrieved sentences are ranked by the conditions described below. Conditions are described in order of priority. If there is more than one sentence having the highest score under these conditions, the most similar sentence is selected randomly.

C1: # of identical words in focus area.
C2: # of synonymous words in focus area.
C3: # of identical words in non-focus area.
C4: # of synonymous words in non-focus area.
C5: # of common function words.
C6: # of different function words.
(the fewer, the higher priority)

Figure 4 shows an example of conditions for ranking. Content words in a focus area of the input sentence are underlined and function words are written in italic.

Input
gaishutsu *shi teiru* aida *ni*, (While I was out), kaban *o* nusuma *re mashi ta* (my baggage was stolen.)

Meaning-equivalent Sentence
baggu *o* nusuma *re ta* (My bag was stolen).

C1	nusumu[5]	1
C2	(kaban = baggu)	1
C3	-	0
C4	-	0
C5	*o, re, ta*	3
C6	*suru, teiru, ni, masu*	4

Figure 4: Example of Conditions for Ranking

5 Modification

The sentence with the highest score among the retrieved meaning-equivalent sentences and its translation are taken. If the retrieved sentence has a synonymous word with the input sentence, the synonymous word in the translation of the retrieved sentence is replaced by the translation of the corresponding word in the input sentence.

Figure 5 shows the replacement of synonymous words in the translation of the retrieved sentence.

[2]Japanese content words are written in sans serif style and Japanese function words in *italic* style. Space characters are inserted into word boundaries in Japanese texts.

[3]The value "others" in all modality/tense groups is omitted.

[5]Words are converted to base form.

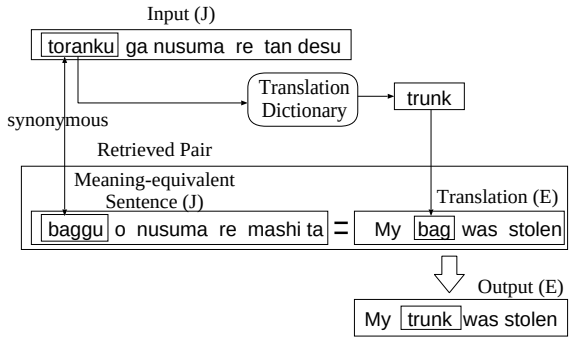

Figure 5: Replacement of Synonymous Words

The sentence "baggu *o* nusuma *re mashi ta*" is retrieved as the most meaning-equivalent sentence of the input "toranku *ga* nusuma *re tan desu*." The word "baggu"(bag) in the retrieved sentence and the word "toranku"(trunk) in the input are synonymous. Therefore, the translation of the retrieved sentence "My bag was stolen" is modified by replacing the word "bag" with "trunk," and the modified translation becomes the output. In this process, the word alignment between the meaning-equivalent sentence and its translation is automatically determined based on a translation dictionary.

6 Experiment

6.1 Test Data

We used a bilingual corpus of travel conversation, which has Japanese sentences and their English translations (Takezawa et al., 2002). This corpus was sentence-aligned, and a morphological analysis was done on both languages by our morphological analysis tools. The bilingual corpus was divided into example data (Example) and test data (Concise) by extracting test data randomly from the whole set of data.

In addition to this, we used a conversational speech corpus for another set of test data (Takezawa, 1999). This corpus contains dialogs between a traveler and a hotel receptionist. It is used to test the robustness against styles. We call this test corpus "Conversational."

We use sentences including more than one content word among the three corpora. The statistics of the three corpora are shown in Table 4.

The thesaurus used in the experiment was "Kadokawa-Ruigo-Jisho" (Ohno and Hamanishi, 1984). Each word has a semantic code consisting of three digits, that is, this thesaurus has three hierarchies. We defined "synonymous" words as sharing exact semantic codes.

Corpus	# of Sentences	Average Length
Example	92,397	7.4
Concise	1,588	6.6
Conversational	800	10.1

Table 4: Statistics of the Corpora

6.2 Compared Methods for Meaning-equivalent Sentence Retrieval

We use two retrieval methods to show the characteristic of the proposed method. The first method (Method-1) adopts "strict" retrieval, which does not allow missing words in input. The method takes function words into account on retrieval. This method corresponds to the conventional EBMT method. The second method (Method-2) adopts "rough" retrieval, which does allow missing words in input, but still takes function words into account. The translation process in these two methods and proposed method is the same.

6.3 Accuracy of Meaning-equivalent Sentence Retrieval

Evaluation was carried out by judging whether the retrieved sentences are meaning-equivalent to the input sentences. The sentences were marked manually as meaning-equivalent or not by a Japanese native-speaker. Figure 6 shows the retrieval accuracy of the three methods with the concise and conversational style data. Retrieval accuracy is defined as the ratio of the number of correctly equivalent sentences to that of the total input sentences. The input sentences are classified into four types by their word length.

The performance of Method-1 reflects the narrow coverage and style-dependency of conventional EBMT. The longer that the input sentences become, the more steeply its performance degrades in both styles. The method can retrieve no similar sentence for input sentences longer than eleven words in conversational style.

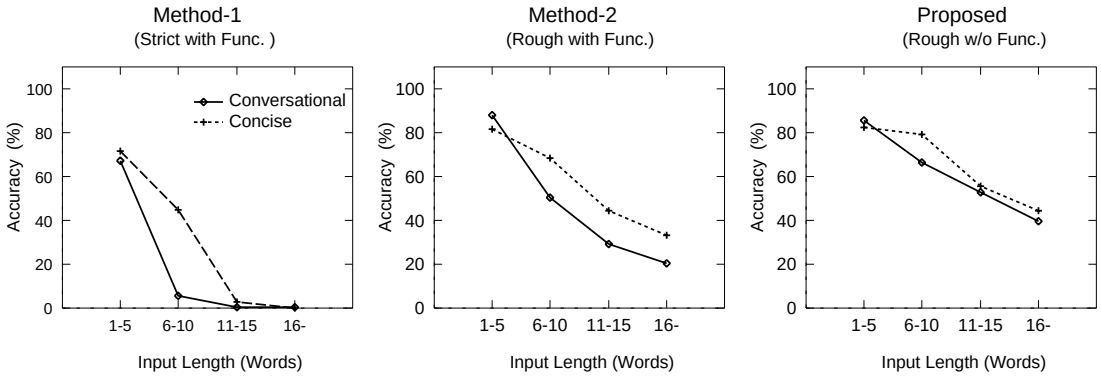

Figure 6: Retrieval Accuracy

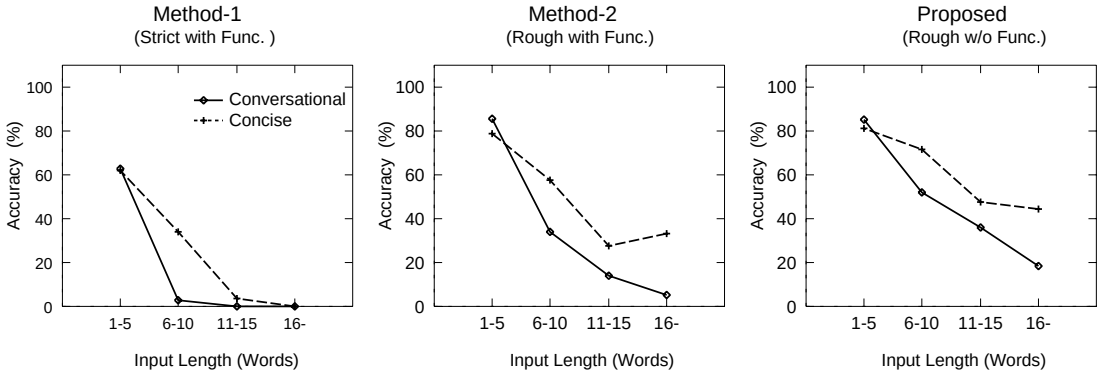

Figure 7: Translation Accuracy

Method-2 adopts a "rough" strategy in retrieval. It attains higher accuracy than Method-1, especially with longer input sentences. This indicates the robustness of the rough retrieval strategy to longer input sentences. However, the method still has an accuracy difference of about 15% between the two styles.

The accuracy of the proposed method is better than that of Method-2, especially in conversational style. The accuracy difference in longer input sentences becomes smaller (about 4%) than that of Method-2. This indicates the robustness of the proposed method to the differences between the two styles.

6.4 Translation Accuracy

Translation accuracy was judged by an English native-speaker. It is defined as the ratio of the number of roughly appropriate translations to that of the total input sentences. Roughly appropriate translations correspond to translations of meaning-equivalent sentences. Figure 7 shows the translation accuracy of the three methods with the concise and conversational style data by input length. As done with retrieval accuracy, the translation accuracy from the proposed method was improved in both long input sentences and conversational styles.

Table 5 shows the overall accuracy for all sentences with conversational data for retrieval accuracy and translation accuracy. The accuracy drop between the retrieval and translation of the rough methods (Method-2 and Proposed) is much larger than that of the strict method (Method-1). One reason for this larger drop is that a context discrepancy between the input sentence and the translation of a meaning-equivalent sentence occurs in the rough methods. This is because unimportant information, which is ignored in rough retrieval methods, has the effect of avoiding the retrieval of sentences having a different context from that of the input sentence.

Method	Retrieval Accuracy (%)	Translation Accuracy (%)
Method-1	25.3%	24.2%
Method-2	54.2%	42.5%
Proposed	63.6%	50.7%

Table 5: Overall Accuracy with Conversational Data

However, retrieval relying on unimportant information degrades total translation accuracy as shown in Table 5. In order to reduce the translation accuracy drop in the rough methods, it is effective to introduce contextual information, such as the scene of the utterance and the type of speaker, in the retrieval process (Yamada et al., 2000).

7 Conclusion

In this paper, we proposed example-based rough translation for S2ST. It aims not at exact translation with narrow coverage but at rough translation with wide coverage. For S2ST, we assume that this translation strategy is sufficiently useful.

Rough translation is based on meaning-equivalent sentences that have the same main meaning as the input sentence despite missing some unimportant information. The retrieval of meaning-equivalent sentences is based on content words, modality, and tense. This strategy of rough translation brings robustness to the input length and the style differences between input sentences and the example corpus. An experiment on travel conversation demonstrated these advantages.

Most MT systems aim to achieve exact translation, but unfortunately they often output bad or no translation for long conversational speeches. Rough translation achieves robustness in translating such input sentences. This method compensates for the shortcomings of conventional MT and makes S2ST technology more practical.

References

R. D. Brown. 2000. Automated generalization of translation examples. In *Proc. of the 18th International Conference on Computational Linguistics (COLING-2000)*.

M. Carl. 1999. Inducing translation templates for example-based machine translation. In *Proc. of the Machine Translation Summit VII*, pages 250–258.

W. B. Frakes and R. Baeza-Yates, editors. 1992. *Information Retrieval Data Structures & Algorithms*. Prentice Hall.

D. Jurafsky and J. H. Martin, editors. 2000. *Speech and Language Processing*. Prentice Hall.

G. Lazzari. 2002. The V1 framework program in Europe: Some thoughts about speech to speech translation research. In *Proc. of 40th ACL Workshop on Speech-to-Speech Translation*, pages 129–135.

M. Nagao. 1981. A framework of a mechanical translation between Japanese and English by analogy principle. In *Artificial and Human Intelligence*, pages 173–180.

S. Ohno and M. Hamanishi, editors. 1984. *Ruigo-Shin-Jiten*. Kadokawa. (in Japanese).

E. Sumita. 2001. Example-based machine translation using DP-matching between work sequences. In *Proc. of the ACL 2001 Workshop on Data-Driven Methods in Machine Translation*, pages 1–8.

T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, and S. Yamamoto. 2002. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *Proc. of the 3rd LREC*, pages 147–152.

T. Takezawa. 1999. Building a bilingual travel conversation database for speech translation research. In *Proc. of the 2nd international workshop on East-Asian resources and evaluation conference on language resources and evaluation*, pages 17–20.

W. Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer.

Alex Waibel. 1996. Interactive translation of conversational speech. *IEEE Computer*, 29(7):41–48.

S. Yamada, E. Sumita, and H. Kashioka. 2000. Translation using information on dialogue participants. In *Proc. of the ANLP-NAACL2000*, pages 37–43.

S. Yamamoto. 2000. Toward speech communications beyond language barrier - research of spoken language translation technologies at ATR -. In *Proc. of International Conference on Spoken Language Processing (ICSLP)*, volume 4, pages 406–411.