Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Intonational Phrasing in a Third Language – The Production of German by Cantonese-English Bilingual Learners

Yanjiao Zhu, Peggy Pik Ki Mok

The Chinese University of Hong Kong
zhuyanjiao33@gmail.com, peggymok@cuhk.edu.hk

1. Abstract

This study looks at intonational phrasing patterns in read speech produced by third language (L3) learners of German who speak Cantonese as the first language (L1) and English as the second language (L2). Acoustic analyses of recordings from 15 L3 learners and 10 native German speakers revealed that intonational phrasing in L3 German was different from that of natives in that 1) L3 learners produced shorter intonational phrases (IP), a few of which were semantically or syntactically incomplete. 2) IP boundary in L3 speech was mainly realized as pause and pitch reset, whereas IP boundary realization in native production was more variegated. 3) Learners used low boundary tones in both continuation and finality statements, while natives adopted high/mid boundary tones for continuation. The study gains insights from and extends the recently developed L2 Intonation Learning theory and offers a multidimensional explanation for phonological acquisition from a third language acquisition perspective.

2. Introduction

Intonational phrase (IP), sometimes referred to as breath group, tone unit, or phonological phrase, is a basic unit in prosody that conveys one piece of information. What is interesting about IP is that it represents an interface of syntax, phonology and discourse. Phonologically, an IP has a complete intonation contour containing at least one nuclear accent and ending with a boundary tone [1]. Semantically, an IP must satisfy the Sense Unit Condition of modifier–head or argument–head relation [2]. Syntactical constituents such as clauses and phrases are also found to partially overlap with IP boundaries [3]. In discourse, tonal contours in IP convey various attitudes in a language-specific way [4], and IP boundary placement serves a disambiguation function [5]. Proper phrasing that chunks the meaning into sense units is vital to achieve communicational purposes [6], whereas a misplacement of IP boundary or an erroneous phonetic implementation of IP not only contributes to foreign accent but also hinders speech intelligibility. Notably, the production of intonational phrases is notoriously hard even for experienced non-native speakers. There have been impressionistic observations describing English IPs produced by Chinese or Cantonese learners as "flat and boring" [7] and those produced by German learners as impolite [8], as well as findings revealing that English speakers divided utterances into too many IPs and applied an overly dynamic pitch movement in L2 German production [9].

Despite sporadic relevant studies in L2 acquisition, until today, no attempt has been made to investigate L3 intonation acquisition, which motivates the present study, aiming at presenting intonational phrasing characteristics of L3 German narrations produced by learners speaking Cantonese as L1 and English as L2. Different from German and English that use pitch height and movement in intonation, Cantonese mainly uses sentence final particles and strictly localized boundary tone for questions [10]. The contrast in L1 and L3 and similarity in L2 and L3 in intonation made it possible to look into the source of transfer in third language acquisition. To further interpret L3 production, this study adopts the framework of L2 Intonation Learning theory (LILt) [11]. According to LILt, crosslinguistic comparison could be made from four dimensions: 1. *Systemic/phonological dimension* is concerned with the distribution of categorical elements (e.g. pitch accents, boundary phenomena and prosodic words) in languages. 2. *Realizational dimension* compares the phonetic realization of such elements in different languages. For instance, it explains cross-linguistic variation in the use of pitch scale, pitch contour and the alignment of nuclear peak. 3. *Semantic dimension* is related to the use of phonological elements to convey meaning. For example, focus can be signaled by pitch accent in some languages but by word order in others. 4. *Frequency dimension* describes the occurrence of phonological primitives, as languages may have the same inventory of pitch accent types but differ in the frequency of using each type.

3. Methods

3.1. Speakers

Non-native acoustic data were taken from 15 undergraduates at the Chinese University of Hong Kong (7 male, 8 female) aged between 20 and 26 (mean 22.3. They were native Cantonese speakers highly proficient in L2 English whose German proficiency was B1 (intermediate) level in European Framework. The subjects were paid for their participation. To set reference, the study extracted native readings of the same material from the *Neuerhebung Corpus* developed by REDE-project [12]. The native speakers were young adults from Berlin, Dresden, Augsburg, and Neubrandenburg, where the Northern Standard German ("Hochdeutch") is spoken.

3.2. Procedure

The L3 learners were recorded individually reading the German *Nordwind und Sonne* ("North wind and the Sun") passage in a quiet room. Reading task rather than spontaneous speaking task was used to allow for clearer comparison, because intonational phrasing shows less variability in read speech, while in spontaneous speech it is often influenced by non-phonological factors such as mental speech planning. To ensure fluency of the read speech, subfajects took time to familiarize themselves with the reading material and were allowed to make multiple attempts. Speech data were recorded at 44 kHz and then transferred to a computer for acoustic analysis. The software Praat [13] was used for labeling and measurement. A sound sequence was marked as an intonational phrase when there were a nucleus and a following pause, which were the major two phonetic correlates for IP [14]. F0 and speech rate were measured automatically using Praat scripts. The extracted F0 data for each speaker were normalized using z-scores. For comparison between L3 learners and native speakers, T-tests were done in the analysis of mean lengths of IP, mean numbers of IP, number of each boundary markers, speech rate variations; frequencies of boundary tones were compared by

10.21437/SpeechProsody.2016-154

Chi-square test; and SS-ANOVA was used to analyze IP intonation.

4. Results

4.1. The division of IP

As shown in Table 1, L3 learners divided utterances into significantly more intonation phrases so that each phrase contained fewer words. The average length of IPs produced by L3 group was shorter than that of natives, $t(22.7) = 2.4$, $p < .05$, and the average number of IPs produced by L3 learners were higher than that of natives, $t(23.0) = -2.5$, $p < .05$.

Table 1. *The mean lengths (in words) and mean numbers of intonational phrases*

	Length of IP	Number of IP
L3	4.55	45.60
Native	5.39	36.30

A high correspondence between intonational phrase and syntactic constituents was substantiated in the current speech data. Table 2 presents the relation between common IP-inducing syntactic structures and the actual IPs produced by over 50% of speakers from their corresponding groups. In both L3 and native speech, sentences and clauses were the most common IP, followed by parenthetical structure, adjuncts and complements. Native speakers showed slightly higher consistency, reaching an agreement of 86.4%, and the figure was 79.4% for learner group.

Table 2. *Grammatical structures of intonational phrases*

Category	Learners	Native
sentence	27.7%	30.3%
subordinate clause	26.8%	28.3%
coordinate clause	9.2%	10.1%
parenthetical structure	8.9%	9.1%
adverbial adjunct	4.6%	2.5%
PP adjunct	2.2%	3.5%
PP complement	--	2.5%
Total	79.4%	86.3%

About 7.4% of IPs produced by L3 learners were considered as ungrammatical, for native German speakers did not produce such kind of phrases. One type of non-target IP included mere functional categories such as "dass (that)", "je mehr (the more)", and "der (the)". Subjects produced them in one individual IP with a nuclear stress, whereas natives never assigned stress on such words or produce them as an isolated IP. Another type of ungrammaticality arose from German separable verbs. Such verbs can have prefixes that are separated from the stem and placed at sentence end. For example, the original verb *ausziehen* was divided into *zog* (past form for *ziehen*) and *aus* in the IP *"zog* der Wanderer seinen Mantel *aus* (the walker took off his coat)". L3 learners tended to place an IP boundary after *zog*, resulting in "*zog*// der Wanderer seinen Mantel *aus*". Such truncation of a verb was syntactically illegal and was not found in native speech.

4.2. Boundary markers

Cruttenden [15] has proposed four criteria for phonetic specifications of IP boundary in English, which were pause, pitch reset, final lengthening, and anacrusis. Since the German IP boundary implementation is similar to English [16], the criteria suggested by Cruttenden were used in the present study. Specifically, a silent period over 200ms was considered a perceptible pause; pitch reset was identified by an increase of F0 at IP initial after F0 declination in the previous IP; final lengthening included cases where the last syllable had longer duration than others in that IP; and anacrusis was determined by a shortening of duration and a lowering of F0 of unstressed syllables before the first stressed syllable. Figure 1 shows that pause and pitch reset were the most common ways to mark an IP boundary in both L3 learner and native groups. In contrast, anacrusis and final lengthening appeared less often in L3 production than in native production. There were significant group differences in the frequencies of anacrusis, $t(13.9) = 5.3$, $p < .001$, and final lengthening, $t(15.3) = 2.6$, $p < .05$.

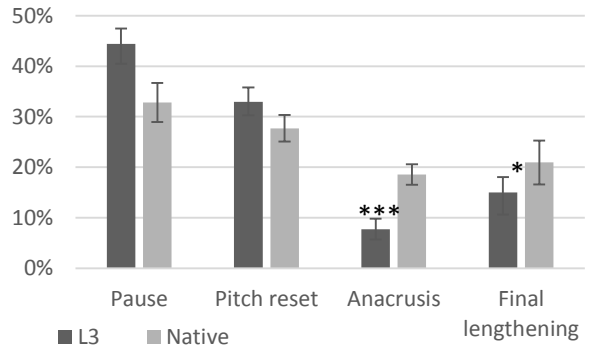

Figure 1: *Frequency distribution of boundary markers in L3 and native productions. (* p < 0.05, ** p < 0.01, ***p < 0.001)*

The low occurrence of anacrusis and final lengthening at L3 IP boundaries is related to inadequacies of using speech tempo variation in L3 speech. Speech rate data were measured with the Praat script developed by de Jong and Wempe [17] which calculates speech rate generated by automatic syllable counting. Compared with natives, the L3 speakers showed lower speech rates (3.18 vs. 2.89), $t(380.0) = 8.8$, $p < .001$ and lower speech rate standard deviations (0.69 vs. 0.84), $t(12.8) = 2.7$, $p < .05$, suggesting that L3 learners spoke slowly with little speech rate variation (Figure 2). Thus the overall lack of speech rate variation as a whole could be one of the reasons for the inadequacy of anacrusis and final lengthening in L3 speech.

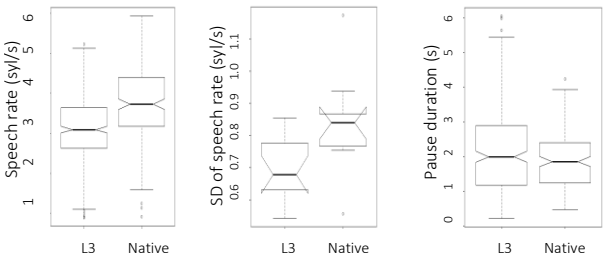

Figure 2: *Boxplots for mean speech rates (left) and standard deviations of speech rates (middle) and pause durations after each IP (right) produced by L3 learners and German natives*

In addtion to the more pauses at boundary, the pauses lasted longer in L3 speech than in native speech (Figure 2). T-test found that the mean durations of all pauses between IPs produced by L3 speakers were significantly longer than IPs produced by native speakers, $t(431.34) = 2.9132$, $p < .01$. This might be a corrollary of the general slower speech rate in non-native production.

4.3. Boundary tones

The boundary tones were categorized into high (H%), mid (M%) and low (L%). Table 3 shows that L% was dominant in both native and L3 speech, in accord with the narrative style of the reading material. This was unsurprising as falling tone was the default type for statements across languages.

Table 3. *Boundary tone inventory produced by L3 and native groups*

	H%	M%	L%
L3	27.40	13.40	59.10
Native	38.70	18.90	52.40

However, a subdivision of statements into continuation and finality statements revealed a divergence of L3 speech from native speech. Table 4 shows the continuation IP and finality IP produced by L3 learners and natives. A Chi-square test of independence was performed to compare the frequency of boundary tone types in L3 and native groups. Learners exhibited a preference for L% in continuation IPs compared to German natives $\chi^2(2)= 34.96$, $p< .001$. As for finality IP, the percentage of each boundary tone did not differ between groups, $\chi^2(2) = 2.54$, $p > .05$.

Table 4. *Frequencies of boundary tones in continuation IP (C) and finality IP (F)*

	L3		Native		Total	
	C	F	C	F	C	F
H%	33	6	41	2	74	8
M%	0	1	14	0	14	1
L%	37	55	7	52	44	107
Total	70	62	62	54	132	116

4.4. F0 analysis of IP

The above difference in boundary tones were further illustrated in Figure 3, presenting the entire intonation contour of IPs produced by an L3 speaker and a native speaker. In the continuation IP "aber je mehr er blies (but the more it blew)", F0 at "blies" was low in the L3 speaker's production (top-left), whereas it was high in the German native speaker's production (top-right). While in finality IP "dass die Sonne von ihnen beiden der stäkere war (that the sun is the stronger of the two)", the F0 at "war" was low in both L3 (bottom-left) and native (bottom-right) productions. This was not an individual but a group phenomenon which can be directly observed from Figure 4 that displays smoothing splines of the intonation contours with 95% Bayesian confidence intervals of the total 6 native and 6 nonnatives continuation IPs (top) and 4 native and 7 nonnative finality IPs (bottom) with the same lexical contents as those in Figure 3. It is clearly seen that the native group raised the pitch while L3 group lowered the pitch at the end of continuation IP. In contrast, the pitch contours for both L3 and native groups converged on a low pitch in the finality case.

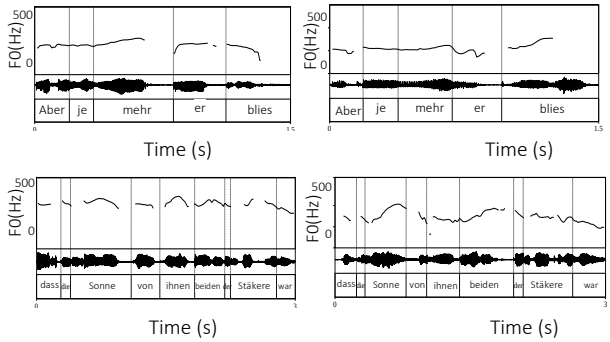

Figure 3: *Representative pitch tracks for continuation IP "aber je meher er blies..." (top-Left: an L3 learner, top-right: a native German), and finality IP "dass die Sonne von ihnen beiden der Stärkere war" (bottom-Left: an L3 learner, bottom-right: a native German)*

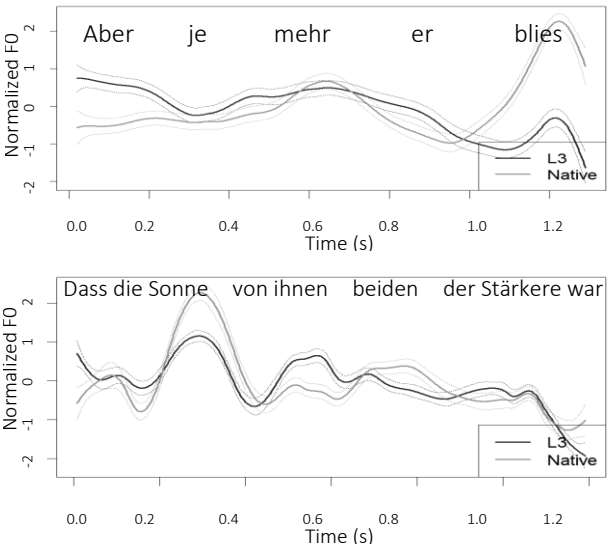

Figure 4: *Smoothing splines estimates and 95% confidence intervals for comparison of intonation contours in the continuation IP "aber je meher er blies" for 6 native and 6 L3 speakers (top) and the finality IP "dass die Sonne von ihnen beiden der Stärkere war" for 4 native and 7 L3 speakers (bottom)*

Within IP, the pitch ranges in L3 speech were smaller than in native speech. F0 maximum and minimun values for each IP were extracted by Xu's ProsodyPro [18], a large-scale Praat script for systematic prosodic analysis. Then F0 ranges of all IPs produced every speaker were calculated. It could be seen from Figure 5 that L3 subjects showed more variablity in F0 range and their F0 range values were higher than those of natives. T-test provided further support for L3 subjects having smaller F0 range than natives, $t (21) = 2.7213$, $p < .05$. This suggested smaller pitch movements within L3 intontational phrases, which could render learners' productions monotonous.

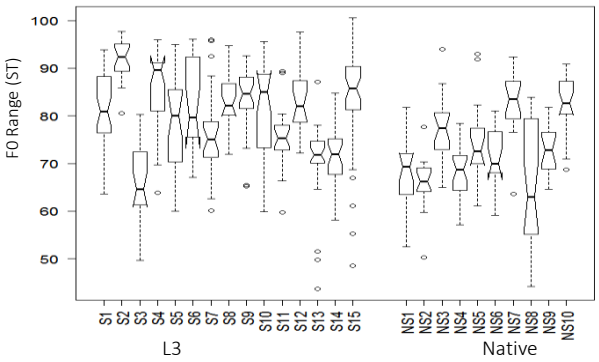

Figure 5: *F0 ranges of IPs produced by L3 and native speakers*

5. Discussion

A satisfactory explanation for non-native speech variation requires a systematic cross-linguistic comparison of native and the target language. Following LILt, the present study analyzed the observed phenomena by comparing the target language (L3 German) with background languages (L1 Cantonese and L2 English) in systemic, realizational, semantic and frequency dimensions.

In systematic dimention, intonational phrasing is quite congruent in learners' target language as well as background languages, being syntactically conditioned in all of three languages concerned. In English and German, IP boundaries coincide with the right edge of root and embedded clauses, sentence adverbials, parallel structure, parentheticals and topicalized elements [3], [19]–[21]. Likewise, an IP sets off a phrase, a clause or a sentence in Cantonese [22]. This similarity of IP inventory and distribution is reflected in speech data where most of L3 and native IPs contained similar syntactic structures. In occasional cases, syntactical structures in German and other languages do not have one-to-one relation, and such structural differences resulted in non-target phrasing. The unique German separable verb prefix occurring at the end of a sentence such as "aus-" does not have correspondence in English and Cantonese, thus incurring erroneous boundary placement. Functional categories in English and German do not stand alone as an IP because they do not receive stress, whereas Cantonese assign equal stress to both lexical and functional categories. Therefore, producing the single relative pronoun "dass" with stress in a single IP should originate from Cantonese stress pattern.

The realizational dimension explains the phonetic specifications of IP boundary and the intonation within IP. Both German and English realize IP boundary mainly by pause, pitch reset, final lengthening, and anacrusis [15] [16]. In comparison, final lengthening and anacrusis are not realized in Cantonese IP boundary, because Cantonese mainly uses pitch span resetting and pausing to mark IP boundary [22]. Therefore, although L3 learners applied all of the four cues in boundary realization, final lengthening and anacrusis did not occur as often as in native production, for those two cues are absent from their native language Cantonese. In addition, pitch movement is the main phonetic cue for intonation in English and German, whereas pitch movement has much less intonational load in Cantonese, a language that expresses phrase edges by boundary tones and particles. Due to the lack of pitch movement for intonation in L1, the third language speakers did not adopt enough falls and rises in German IP, resulting in smaller pitch ranges within each IP.

The choice of boundary tones in expressing meanings falls into the semantic domain. Whereas rising [23]–[25] or level tone [4] are often used in German statements, falling tone is the default in English [16] and Cantonese [22] statements. This difference was evidenced in the L3 data where superfluous low boundary tones in continuation statements were observed. Apart from meaning deviance, meaning integrity was another problem in L3 speech. According to the Sense Unit Condition, an IP should form a complete sense unit, whose components are either in modifier–head or in argument–head relation [2]. IPs containing only functional categories violated Sense Unit Condition, so they were nonexistent in native utterances. In contrast, those IPs were produced by L3 learners, who tended to partition long sentences into more and shorter phrases, oftentimes yielding ungrammaticality or incompleteness of information.

Lastly, it is observed that the four phonetic cues were more evenly distributed in native production than in L3 production, while L3 learners leaned towards non-temporal cues to mark IP boundary. This is related to the frequency of using temporal cues in target and background languages. As a typical syllable-timed language, Cantonese seldom uses strong-weak alternation or temporal variation, whereas stress-timed languages like German and English often vary the durations of syllables to realize stress [26]. Hence temporal cues occur more frequently in L2 and L3, but less frequently in L1, and therefore, the infrequent occurrence of anacrusis and final lengthening at IP boundary might be L1-induced.

The four dimensional comparison suggests that L3 learners' intonational phrasing patterns are mainly L1-influenced. In spite of the resemblance English and German intonation and years of training in English, the speakers' native language Cantonese still exerts strong influences on L3 intonation.

6. Conclusion

The present study evaluated the third language intonational phrasing in read speech produced by 15 intermediate learners of German with L1 Cantonese and L2 English. Acoustic analyses comparing L3 learners with native speakers found that L3 speakers produced more and shorter IPs, a small portion of which did not constitute a full sense group. Subjects did not employ flexible phonetic cues at IP boundaries nor abundant pitch dynamics to the same extent as natives did. In L3 intonation, no distinction was made between IPs with continuation and with finality implications, the latter being marked by a high or level boundary tone in native German production. The comparison between background languages and the target language suggests a strong influence from the first language Cantonese in L3 production of IP. By adopting LILt, the present study offers a tentative theoretical explanation for L3 production and hopes to shed light on L3 phonological acquisition.

7. Acknowledgements

This work is partially supported by the National Social Science Fund of China (10CYY009) and the Major Program for the National Social Science Fund of China (13&ZD189).

8. References

[1] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook*, vol. 3. pp. 255–309, 1986.

[2] E. O. Selkirk, *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, 1984.

[3] M. Nespor and I. Vogel, *Prosodic Phonology*. Dordrecht: Fori., 1986.

[4] D. J. Hirst and A. Di Cristo, *Intonation systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998.

[5] P. Tench, “Intonation and the differentiation of syntactic patterns in English and German,” *Int. J. Appl. Linguist.*, vol. 6, no. 2, pp. 223–256, 1996.
[6] M. A. K. Halliday, *Intonation and Grammar in British English*. The Hague: Mouton, 1967.
[7] Y. W. A. Chan, D. C. S. Li, A. Y. W. Chan, and D. C. S. Li, “English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners’ English Pronunciation Problems,” *Lang. Cult. Curric.*, vol. 13, no. 1, pp. 67–85, Jan. 2000.
[8] W. G. Moulton, *The Sounds of English and German*. Chicago: The University of Chicago Press, 1962.
[9] U. Gut, *Non-native Speech*. Frankfurt am Main: Peter Lang GmbH, 2009.
[10] V. Yip and S. Matthews, *Cantonese: a Comprehensive Grammar.*, 2nd editio. London: Routledge, 2011.
[11] I. Mennen, “Beyond segments: towards a L2 intonation learning theory,” in *Prosody and Language in Contact L2 Acquisition, Attrition and Languages in Multilingual Situations*, E. Delais-roussarie, M. Avanzi, and S. Herment, Eds. Berlin,Heidelberg: Springer Verlag, 2015, pp. 171–188.
[12] J. E. Schmidt, J. Herrgen, and R. Kehrein, “Regionalsprache.de (REDE). Forschungsplattform zu den modernen Regionalsprachen des Deutschen.” Forschungszentrum Deutscher Sprachatlas, Marburg, 2008.
[13] D. Boersma, P., & Weenink, “Praat: doing phonetics by computer.” 2015.
[14] J. B. Pierrehumbert and J. Hirschberg, “The meaning of intonational contours in the interpretation of discourse,” in *Intentions in Communication*, Cambridge MA: MIT Press, 1990, pp. 271–311.
[15] A. Cruttenden, *Intonation*, Second edi. New York: Cambridge University Press, 1997.
[16] E. Grabe, “Intonational Phonology: English and German,” Max-Planck-Institut for Psycholinguistics and University of Nijmegen, 1998.
[17] N. H. de Jong and T. Wempe, “Praat script to detect syllable nuclei and measure speech rate automatically.,” *Behav. Res. Methods*, vol. 41, no. 2, pp. 385–390, 2009.
[18] Y. Xu, “ProsodyPro - A tool for large-scale systematic prosody analysis,” in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP2013)*, 2013, pp. 7–10.
[19] B. T. Downing, “Syntactic structure and phonological phrasing in English,” University of Texas at Austin, 1970.
[20] H. Truckenbrodt, “A short report on intonation phrase boundaries in German,” *Linguist. Berichte*, vol. 203, no. October 2004, pp. 273–296, 2005.
[21] J. C. Wells, *English Intonation: An Introduction*. New York: Cambridge University Press, 2006.
[22] Choi-Yeung-Chang Flynn, *Intonation in Cantonese*. Munich: LINCOM GmbH, 2003.
[23] C. Grover, D. G. Jamieson, and M. B. Dobrovolsky, “Intonation in English, French and German: Perception and Production,” *Lang. Speech*, vol. 30, pp. 277–295, 1987.
[24] J. Trim, “Some contrastive intonated features of British English and German,” in *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel’s 60th Birthday*, J. Klegraf and D. Nehls, Eds. Heidelberg: Julius Groos, 1988, pp. 235–349.
[25] E. Delattre, P. Poenack, and C. Olsen, “Some Characteristics of German Intonation for the Expression of Continuation and Finality,” *Phonetica*, vol. 13, pp. 134–161, 1965.
[26] P. Mok, “On the syllable-timing of Cantonese and Beijing Mandarin,” *Chinese J. Phonetics*, vol. 2, pp. 148–155, 2009.