



## Dynamic Assessment during Suprasegmental Training with Mobile CAPT

Veranika Mikhailava<sup>1</sup>, John Blake<sup>1</sup>, Evgeny Pyshkin<sup>1</sup>, Natalia Bogach<sup>2</sup>, Sergey Chernonog<sup>2</sup>, Artem Zhuikov<sup>2</sup>, Maria Lesnichaya<sup>2</sup>, Iurii Lezhenin<sup>2,3</sup>, Roman Svechnikov<sup>2</sup>

<sup>1</sup>The University of Aizu, Aizu-Wakamatsu, Japan

<sup>2</sup>Peter the Great State Polytechnic University, St. Petersburg, Russia

<sup>3</sup>Speech Technology Center Ltd, St. Petersburg, Russia

pyshe@u-aizu.ac.jp, bogach@kspt.icc.spbstu.ru

### Abstract

This paper reports the results of a small-scale longitudinal study on the use of *StudyIntonation*, a computer-assisted pronunciation teaching environment. *StudyIntonation* aims to scaffold learners through their zone of proximal development by drawing eclectically on concepts, such as Vygotskian socio-cultural theory, dynamic assessment and second language development. Learners perform shadowing tasks, aiming to replicate the suprasegmental prosodic aspects of model sentences. The pitch curves of the model and user attempts are displayed to help learners see their progress. We observed a group of learners who performed shadowing tasks in *StudyIntonation* for 24 months. The resultant corpus comprises 1050 speech records labelled with orthographic transcript, pitch readings, and similarity metrics. Longitudinal and microgenetic analysis of L2 pronunciation development was conducted on this dataset. Prosodic synchronization between speakers as well as longitudinal pronunciation assessment allows for quantitative evaluation by means of a dynamical modeling technique of cross-recurrence quantification analysis (CRQA). We located the zone of proximal development of each learner, where she/he reveals an increased responsiveness to input audiovisual stimuli, through the oscillations of pitch similarity metrics of dynamic time warping and CRQA. The rates of cross-recurrence between the model and learner are helpful synchronization indicators and performance predictors.

**Index Terms:** second language development (SLD), computer-assisted pronunciation training (CAPT), socio-cultural theory (SCT), dynamic assessment (DA), actual development zone (ADZ), zone of proximal development (ZPD), scaffolding, dynamic system theory (DST), pitch similarity, dynamic time warping (DTW), cross-recurrence quantification analysis (CRQA)

### 1. Introduction

Dynamic system theory (DST) has gained attention as a holistic foundation of second language development (SLD) theories [1, 2, 3]. Within the DST view, all agents of language interaction are seen as dynamic systems and all the SLD processes, which occur for all language competencies [4, 5] at various timescales, are explored with respect to their dynamics [6]. A plethora of research on language development considers language acquisition as the emergence of language abilities over time and through language use, and not just as a process of acquiring abstract rules [7, 6, 5, 8]. Emergence is understood as spontaneous acquisition of new features and forms as a result of self-organizing interactions of complex system components [9, 10].

The idea of learning as an emergent process was described within socio-cultural theory (SCT) [11, 12, 13], which considers L2 development as a process of social mediation, thus sharing an emphasis on the role of environmental contexts with DST [14, 15]. SCT-informed L2 pedagogy pays much attention to individual development trajectories and operates with such constructs as zone of proximal development (ZPD) [14], scaffolding [16], mediation [17], inner speech [18] and dynamic assessment (DA) [19]. ZPD is region through which learners improve from their actual level to their potential level under guidance and through feedback [2, 14]. DA is understood as a way to infer about learner abilities and to move beyond performance assessment towards understanding of the processes underlying individual learning dynamics [14, 12]. ZPD and scaffolding were articulated in terms of contemporary DST theory in [20, 1, 16] making possible quantitative research with DST instruments and nonlinear time-series techniques [21, 15, 10]. The field of DST for L2 pronunciation teaching and learning is still under-addressed at present [22, 23], but there is evidence that DST applied to L2 pronunciation teaching and longitudinal L2 pronunciation assessment is an insightful tool to explore the individual progress and motivation of learners [24, 25]. Using DST as a theoretical framework might bridge the gap between L2 phonology and pronunciation teaching [26, 27]. Under the assumption that speech is inherently recurrent [28, 29, 30], speech prosodic phenomena allow for quantitative evaluation by means of dynamical modeling technique of cross-recurrence quantification analysis (CRQA) [31, 32].

This research undertakes an attempt to explore L2 suprasegmental teaching and learning as a complex dynamic process featuring variability, self-organization, and emergence. We pivot upon DST and SCT L2 pedagogy concepts and propose how DA could be constructed in *StudyIntonation* [33], a multimodal computer-assisted pronunciation training (CAPT) environment for Standard British English phrasal intonation. We argue that L2 suprasegmental learning based on SCT constructs and instrumented with dynamic model and CRQA sheds light on processes occurring in the course of learner's interaction with a suprasegmental-oriented CAPT. The following questions guided this research:

1. How a dynamic model could be applied to L2 suprasegmental teaching and learning with a CAPT system?
2. How developmental processes of learners in terms of their ZPD could be observed in the course of suprasegmental training?
3. How DA approach based on learners developmental trajectories could add to more individualized CAPT feedback and instruction focus?

The remainder of this paper is structured as follows: Section 2 explains the method of DA built on integrating a dynamic model for Vygotskian ZPD into the CAPT system; Section 3 describes and discusses the experiments leading towards a DA method for CAPT system; while Section 4 summarises the research outcome.

## 2. Methodology

Research informed by DST is defined and shaped by specific questions under consideration, such as the emergence and dynamics of specific language skills [34, 4], but there is still a need for an explicit form of applicability of dynamic models in L2 empirical research designs [35, 36]. We adapted a dynamic model for Vygotskian developmental mechanism from [20, 1, 9] for a narrow context of a CAPT system interaction.

We used the learning content of *StudyIntonation*, i.e. a subset of English phrasal intonation patterns, as the array of external and internal contents [1]; metrics based on dynamic time warping (DTW) [37] and CRQA descriptors as indicators of learning dynamics (in the form of phase shifts, developmental jumps, etc.) [10]. We searched for quantitative indicators of learners' entrance into their ZPD in the course of CAPT system interaction alongside with spotting the specific tasks, which could be of maximum usefulness because of sensitivity, responsiveness and perception increase while being within one's ZPD. Thus, performing assessment and instruction together, we obtain individually tailored DA-based feedback.

### 2.1. Dynamic Model of Vygotskian Developmental Mechanism

Dynamic system current state generates its successive state by a rule of change ("evolution rule", the driver of its change) and thus produces a trajectory in a state space [9]:

$$x_{t+1} = f(x_t) \rightarrow x_{t+2} = f(x_{t+1}) \rightarrow \dots \quad (1)$$

Van Geert [20] incorporates the Vygotsky dialectical mechanism of development into the dynamic model in the following way: performing an action, the system (e.g., a learner) activates a particular content  $c_n$  (a pattern, a skill, a rule, etc.), which is associated with a specific developmental level  $n$  that is responsible for that action. Any event which can be either an action or an experience is a confluence between an internal content  $c_n \in I, I(c_1, c_2, \dots, c_n, \dots)$  and an external content  $c_n \in E, E(c_1, c_2, \dots, c_n, \dots)$  which share the same developmental level index  $n$ .  $I \subseteq E$ , that implies the environment  $E$  is viewed as a potential source of learning and development.

The effect of an action on further development of the system, depends on two contents – the first is the activated content  $c_n$  which represents its actual developmental level  $A_i$ , which Vygotsky names the actual development zone (ADZ); the second content  $c_k, k > n$  is defined by the help or information or feedback resulting from performing the action. This second content defines or specifies the system's potential level  $P_i$ , which is a developmental level corresponding to a set of contents (patterns, skills, rules, etc.) in the array  $I$  that is most sensitive to the effect of experience brought about by the activated content  $c_k$ . This sensitivity to instruction is what Vygotsky contended with his ZPD concept.

The existence of a content that is more sensitive to experience than the others is, according to [1], based on two opposing tendencies that are likely to occur in learning and developing systems: preference for novelty and preference for familiarity.

These tendencies can be expressed in the form of a pair of exponential function [20, 1]:

$$f_{familiar}(i) = ab^{ci} \quad (2)$$

$$f_{novel}(i) = dg^{fi} \quad (3)$$

where  $a, d, f > 0, b, g \in (0, 1), c < 0$ , and  $i$  is the distance between contents in the array  $I$ . Van Geert [1] names the most preferred content for both functions as the one at the cross-section point of  $f_{familiar}(i)$  and  $f_{novel}(i)$ :

$$i = n + \Delta M = \frac{\log\left(\frac{a}{d}\right)}{f \log(g) - c \log(b)} \quad (4)$$

Each time the system has gone through an action/experience the content arrays are updated. The maximal gain occurs at two places: at the content corresponding with the actual output level  $A_i$ , and the balance point of maximum sensitivity to experience which is defined by the equation 4.

The mathematical model for Vygotsky ZPD concept of learning includes two evolution rules for actual and potential developmental levels [20]:

$$A_{i+1} = A_i(1 + R_{A_i} - R_{A_i} \frac{A_i}{P_i}) \quad (5)$$

$$P_{i+1} = P_i(1 + R_{P_i} - R_{P_i} \frac{P_i}{P_k}), \quad (6)$$

where  $R_{A_i}$  – learning rate,  $R_{P_i}$  – teaching rate,  $P_k$  – goal state. The learning and teaching rates update rules are defined as:

$$R_{A_i} = r_A - \left| \frac{P_i}{A_i} - o \right| \alpha (P_k - P_i) \quad (7)$$

$$R_{P_i} = r_P - \left| \frac{P_i}{A_i} - o \right| \beta (P_k - P_i), \quad (8)$$

where  $r_A, r_P$  are constant growth factors,  $o$  sets the optimal  $\frac{P}{A}$  ratio,  $\alpha, \beta$  are damping parameters. This model describes the reciprocal interaction between the principal variables  $A_i$  and  $P_i$  with the support of a set of control variables. The effect of dynamic model is that it allows for observing the learner's gradual transition from one developmental level to another.

### 2.2. CAPT Experimental Platform *StudyIntonation*

DST was shown to be sound and beneficial in digital environments for language learning [38, 23], where the importance and contribution of multimodal input was supported by neuroimaging studies on the optimal operation of the human brain in multisensory environments [39]. Interaction with *StudyIntonation* is multimodal, involving listening to a recording, observing a pitch curve and shadowing a model phrase. As a means of feedback the system produces a pitch curve for recorded speech and calculates DTW-based similarity metrics. The learning content of *StudyIntonation* is produced by native speakers, it covers a set of 74 examples of phrasal intonation patterns and is structured in 4 groups with respect to various speech situations. This approach is in line with recent SLD research, which highlights the influence of social factors to language acquisition and argues a holistic, top-down approach as paramount for L2 pronunciation instruction [8]. The most CAPT resources are still in need for accurate instructive feedback and learning strategies [6, 40]. As L2 suprasegmental teaching and learning may be understood as dynamic process, we searched to improve the instruction and feedback mechanism by mapping CAPT context to a dynamic model (Table 1).

### 2.3. CAPT Environment as a Dynamic Model with DTW and CRQA Metrics as Developmental Descriptors

CAPT system	Dynamic model
Learner (narrow context)	Internal array of contents, skills, etc. $I(c_n)$
Courseware (source of experience)	External array of contents, events, patterns, etc. $E(c_n)$
A specific pattern (task)	Familiar content in ADZ $c_n$ Novel content in ZPD $c_k$
Performance metrics (DTW, CRQA)	Developmental indicators

Table 1: Correspondence between DST Concepts and Components of CAPT System for Suprasegmental Learning

The choice of DTW and CRQA as developmental indicators is motivated by the fact, that when taken longitudinally, they reflect the changes or represent properties of intonation acquisition and model/learner prosodic synchronization. CRQA metrics discriminating and descriptive ability for various L2 DST-based research, speech synchronisation and emotion recognition tasks was repeatedly demonstrated (see, e.g. [10, 32]), while DTW is known as a conventional measure of pitch curve similarity [37].

The interpretation of DTW-based performance metrics in terms of dynamic model of ZPD relies upon the admission that a specific DTW-based metric depends upon the familiarity/novelty of a specific content and may be understood as a familiarity/novelty measure of a specific experience, thus a DTW-based metric could be used as a visible indicator of a specific content activation. DTW metric variability, thus, should signal learner’s transition from one developmental level to another [4] and produce unimodal and bimodal curves. The transition is detected by the emergence of a second peak yielding a bimodal pattern. After the transition, the first peak disappears and the original unimodal pattern at a higher level is restored [1].

## 3. Results and Discussion

### 3.1. StudyIntonation Dataset Collection

A dataset of 1050 records of 1 British English native speaker and 5 non-native speakers with L1 Russian doing *StudyIntonation* tasks was collected. A group of learners performed a shadowing task during one-hour learning sessions occurring intermittently over a 24-month period.

The whole observation time was split into task-wise and session-wise timescales:

1. Two successive attempts of one task (30 sec);
2. All attempts for a specific task (15 min);
3. All attempts for all tasks covered within a learning session (45-60 min);
4. All attempts for a specific task within the observation period (24 months); and
5. All attempts for all tasks within the observation period (24 months).

Since pronunciation dynamics demonstrates short-term and acute shifts [41] the most significant developmental effects were obtained for scales 1 and 2 (Fig. 1).

Each record in the dataset bears the following markup:

1. Subjective expert decision (SED) whether learner’s attempt is good or not (binary, 0 or 1);
2. DTW metric of model and learner pitch similarity (DTW); and
3. CRQA metrics: cross-recurrence rate (RR), percentage of determinism (DET), average diagonal length (AVG\_DIAG), etc.[31].

The embedding dimension [31] for CRQA metrics  $d \in [2, 4]$  for pitch contours was obtained by the False Nearest Neighbors algorithm (FNN) [42]. This value complies with DST-based speech  $f_0$  extraction algorithms, (e.g. [29, 30]), where the number of embedding dimensions for  $f_0$  was shown to be within an interval of 2 to 5. CRQA metrics were calculated for embedding dimension,  $d = 3$  and point proximity radius,  $\epsilon = 2$ .

### 3.2. DTW and CRQA as Joint Developmental Descriptors

To explore whether using DTW and CRQA metrics together add to the learner performance assessment more than when used separately, we trained logistic binary classifiers to discriminate learner attempts from the *StudyIntonation* dataset and examined DTW and CRQA feature importance (Table 2). Logistic regression classifiers were trained using DTW together with various CRQA feature subsets as a part of the input. Table 2 shows classifier accuracy for DTW; DTW, RR; and RR, DET feature sets, where a combination of DTW and RR shows the best discrimination ability.

- DTW-based classification had the lowest accuracy of 0.60.
- CRQA-based classifier had relatively poor accuracy of 0.63.
- DTW+RR-based classifier achieved better accuracy of 0.73.
- Other CRQA features, when included into the feature set, confused the logistic classifier.

Feature set	Classifier metrics ( $p$ – precision, $r$ – recall)	Accuracy (F1)
DTW	$p_0: 0.65$ $r_0: 0.56$ $p_1: 0.56$ $r_1: 0.65$	0.60
RR, DET	$p_0: 0.70$ $r_0: 0.61$ $p_1: 0.61$ $r_1: 0.69$	0.65
DTW, RR	$p_0: 0.75$ $r_0: 0.68$ $p_1: 0.67$ $r_1: 0.73$	0.71

Table 2: DTW and CRQA Feature Sets

### 3.3. Transition to ZPD. DA within CAPT Interaction

In [5, 4, 13] it was shown that learners’ performance does not increase linearly, but passes through periods of progression and regression alternatively. These are not isolated jumps, but the stages of a continuous developmental process; and each individual learner demonstrated unique patterns of this developmental trajectory.

We located a transition and, thus, a ZPD of each learner by oscillations of DTW and observed an increased responsiveness to input audiovisual stimuli, manifesting itself by the occurrence of low DTW. The rate of cross-recurrence RR in ADZ tends to grow smoothly alongside with a decrease of DTW (Fig.

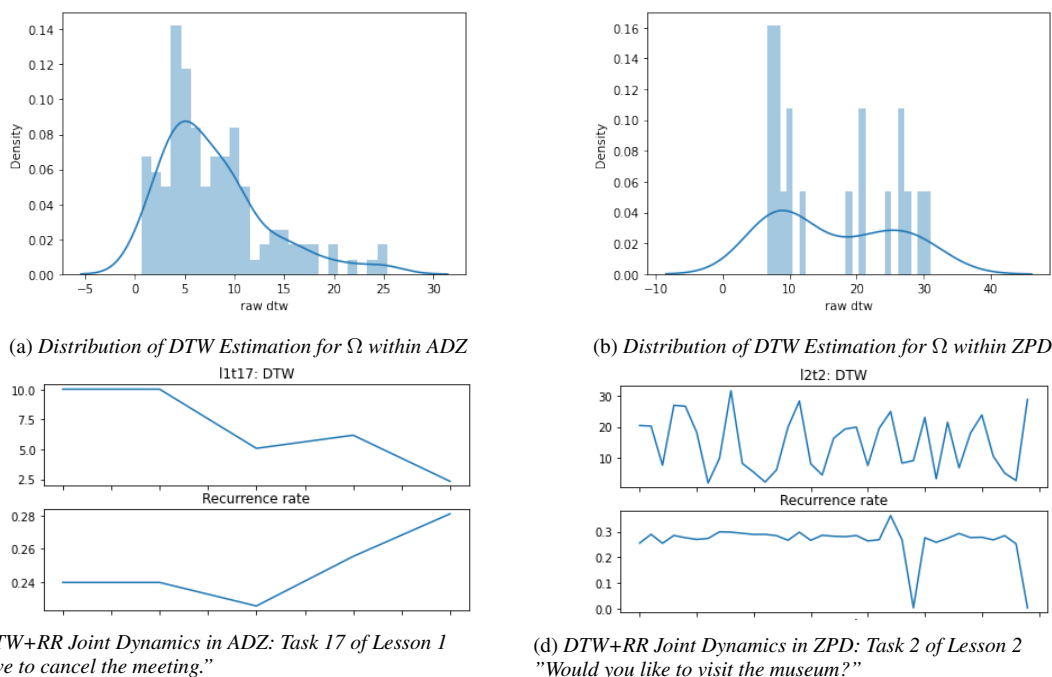


Figure 1: Microgenetic effects at 15 min timescale, which covers all attempts for a specific task within one training session of one learner X. Transition to ZPD is located by oscillation of DTW between two levels

1a,1c). During the state of transition, DTW oscillates between two levels, while RR dynamics may be either indifferent or oscillate randomly (Fig. 1b, 1d). We fixed the specific types of tasks, where such an oscillation started to occur and where learner’s efforts should be directed to (Table 3). Although not the focus of this research, a notable finding was that a non-native speaker was, at times, able to replicate the model more accurately than the native speaker. It appears that possessing “a good ear” may be a more important determinant of success than mother tongue.

ADZ
L1T17: <i>I had to cancel the meeting.</i> L1T18: <i>I'd really appreciate that.</i>
ZPD
L1T19: <i>Are you going for lunch now?</i> L2T1: <i>Would you like to join me for dinner?</i> L2T2: <i>Would you like to visit the museum?</i> L1T25: <i>I'm glad, that's really kind of you, thank you!</i>

Table 3: Example of DA-Driven Task Assignment for StudyIntonation in Accordance with Learner’s ADZ and ZPD location

#### 4. Conclusion

Teaching individual segmental and suprasegmental features can positively influence the global construct of L2 pronunciation proficiency [43, 44]. Maximum sensitivity to particular contents of developmental levels means that experiences at those levels yields a maximal effect [1]. The major outcome of this research is how to perform DA during the process of phrasal intonation teaching with a CAPT system and how to determine learners’ movement from one developmental level to another.

While working through ADZ, DTW metrics are either immediately low or rapidly and steadily converge to small values. A good rising edge of RR is present, which indicates that two phonological systems are synchronizing with each other. When transition to a new level (ZPD entrance) is approaching, there is a group of tasks where DTW is high, but immediately after instruction there is a short effect of prosodic memory, which results in a low DTW metric for one attempt. The student produces a good result, but cannot hold this effect longer. This variability signals a maximum sensitivity to the instruction and experience. It is necessary to spot the type of tasks, where oscillations occur, specific for each learner, and direct the focus of efforts there. In the example in Fig. 1, these are longer interrogative and exclamatory sentences (Table 3). A distant ZPD is formed by contents where learners can not yet access the model and all indicators are unstable; but step by step, as a consequence of teaching in the ZPD, learners become more familiar with contents that are ahead of their current actual developmental level [9, 4, 6, 34].

#### 5. Acknowledgements

The project was partially supported by the Japan Society for the Promotion of Science (JSPS), grant 20K00838.

#### 6. References

- [1] P. Van Geert, “A dynamic systems model of basic developmental mechanisms: Piaget, vygotsky, and beyond.” *Psychological review*, vol. 105, no. 4, p. 634, 1998.
- [2] D. Larsen-Freeman, “Chaos/complexity science and second language acquisition,” *Applied linguistics*, vol. 18, no. 2, pp. 141–165, 1997.
- [3] P. Hiver, A. H. Al-Hoorie, and R. Evans, “Complex dynamic systems theory in language learning: A scoping review of 25 years

- of research,” *Studies in Second Language Acquisition*, pp. 1–29, 2021.
- [4] M. Verspoor and K. de Bot, “Measures of variability in transitional phases in second language development,” *International Review of Applied Linguistics in Language Teaching*, 2021.
- [5] P. Chang and L. J. Zhang, “A cdst perspective on variability in foreign language learners’ listening development,” *Frontiers in Psychology*, vol. 12, p. 21, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsyg.2021.601962>
- [6] D. Larsen-Freeman, “On language learner agency: A complex dynamic systems theory perspective,” *The Modern Language Journal*, vol. 103, pp. 61–79, 2019.
- [7] —, “Saying what we mean: Making a case for ‘language acquisition’ to become ‘language development’,” *Language Teaching*, vol. 48, no. 4, pp. 491–505, 2015.
- [8] D. LaScotte, C. Meyers, and E. Tarone, “Voice and mirroring in sla: Top-down pedagogy for l2 pronunciation instruction,” *REL C Journal*, vol. 52, no. 1, pp. 144–154, 2021.
- [9] P. L. van Geert, “Dynamic systems, process and development,” *Human development*, vol. 63, no. 3–4, pp. 153–179, 2019.
- [10] D. R. Evans, “Bifurcations, fractals, and non-linearity in second language development: A complex dynamic systems perspective,” Ph.D. dissertation, State University of New York at Buffalo, 2019.
- [11] L. S. Vygotsky, *Thought and language*. MIT press, 2012.
- [12] M. Koopmans, “Education is a complex dynamical system: Challenges for research,” *The Journal of Experimental Education*, vol. 88, no. 3, pp. 358–374, 2020.
- [13] J. Hodgetts, *Pronunciation Instruction in English for Academic Purposes: An Investigation of Attitudes, Beliefs and Practices*. Springer, 2020.
- [14] J. P. Lantolf, L. Kurtz, and O. Kisselev, “Understanding the revolutionary character of l2 development in the zpd: Why levels of mediation matter,” *Language and Sociocultural Theory*, vol. 3, no. 2, pp. 153–171, 2016.
- [15] K. M. Bragg, “Conversational movement dynamics and nonverbal indicators of second language development: A microgenetic approach,” Ph.D. dissertation, University of Nevada, 2018.
- [16] P. van Geert and H. Steenbeek, “The dynamics of scaffolding,” *New ideas in Psychology*, vol. 23, no. 3, pp. 115–128, 2005.
- [17] M. E. Poehner, J. Zhang, and X. Lu, “Computerized dynamic assessment (c-da): Diagnosing l2 development according to learner responsiveness to mediation,” *Language testing*, vol. 32, no. 3, pp. 337–357, 2015.
- [18] B. Tomlinson, “Assisting learners in orchestrating their inner voice for l2 learning,” *Language Teaching Research Quarterly*, vol. 19, pp. 32–47, 2020.
- [19] R. Ableeva, “Dynamic assessment of listening comprehension in second language learning,” Ph.D. dissertation, The Pennsylvania State University, College of the Liberal Arts, 2010.
- [20] P. van Geert, “Vygotskian dynamics of development,” *Human development*, vol. 37, no. 6, pp. 346–365, 1994.
- [21] M. Guevara, R. F. Cox, M. van Dijk, and P. van Geert, “Attractor dynamics of dyadic interaction: A recurrence based analysis,” *Nonlinear Dynamics Psychology and Life Sciences*, vol. 21, no. 3, pp. 289–317, 2017.
- [22] R. M. Lima Jr and U. K. Alves, “A dynamic perspective on l2 pronunciation development: bridging research and communicative teaching practice,” *Revista do GEL*, vol. 16, no. 2, pp. 27–56, 2019.
- [23] A. Bahari, X. Zhang, and Y. Ardasheva, “Establishing a nonlinear dynamic individual-centered language assessment model: a dynamic systems theory approach,” *Interactive Learning Environments*, pp. 1–23, 2021.
- [24] C. L. Nagle, “Assessing the state of the art in longitudinal l2 pronunciation research: Trends and future directions,” *Journal of Second Language Pronunciation*, 2021.
- [25] N. Esteve-Gibert and B. Guellai, “Prosody in the auditory and visual domains: A developmental perspective,” *Frontiers in Psychology*, vol. 9, p. 338, 2018.
- [26] D. Liu and M. Reed, “Exploring the complexity of the l2 intonation system: An acoustic and eye-tracking study,” *Frontiers in Communication*, vol. 6, p. 51, 2021.
- [27] D. Liu, “Prosody transfer failure despite cross-language similarities: Evidence in favor of a complex dynamic system approach in pronunciation teaching,” *Journal of Second Language Pronunciation*, vol. 7, no. 1, pp. 38–61, 2021.
- [28] A. Shahzadi, A. Ahmadyfard, A. Harimi, and K. Yaghmaei, “Speech emotion recognition using nonlinear dynamics features,” *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 23, 2015.
- [29] S. Sood and A. Krishnamurthy, “A robust on-the-fly pitch (otfp) estimation algorithm,” in *Proceedings of the 12th annual ACM international conference on Multimedia*, 2004, pp. 280–283.
- [30] D. E. Terez, “Robust pitch determination using nonlinear state-space embedding,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2002, pp. I–345.
- [31] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5, pp. 237–329, 2007.
- [32] E. Tzinis, G. Paraskevopoulos, C. Baziotis, and A. Potamianos, “Integrating recurrence dynamics for speech emotion recognition,” *Interspeech 2018*, Sep 2018.
- [33] N. Bogach, E. Boitsova, S. Chernonog, A. Lamtev, M. Lesnichaya, I. Lezhenin, A. Novopashenny, R. Svechnikov, D. Tsikach, K. Vasiliev *et al.*, “Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching,” *Electronics*, vol. 10, no. 3, p. 235, 2021.
- [34] W. M. Lowie and M. H. Verspoor, “Individual differences and the ergodicity problem,” *Language Learning*, vol. 69, pp. 184–206, 2019.
- [35] P. Hiver and A. H. Al-Hoorie, “A dynamic ensemble for second language research: Putting complexity theory into practice,” *The Modern Language Journal*, vol. 100, no. 4, pp. 741–756, 2016.
- [36] —, *Research Methods for Complexity Theory in Applied Linguistics*. Multilingual Matters, 2019. [Online]. Available: <https://doi.org/10.21832/9781788925754>
- [37] A. Rilliard, A. Allauzen, and P. Boula\_de\_Mareuil, “Using dynamic time warping to compute prosodic similarity measures,” in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [38] R. Godwin-Jones, “Chasing the butterfly effect: Informal language learning online as a complex system,” *Language Learning & Technology*, vol. 22, no. 2, pp. 8–27, 2018.
- [39] D. M. Hardison, “Multimodal input in second-language speech processing,” *Language Teaching*, vol. 54, no. 2, pp. 206–220, 2021.
- [40] M. C. Pennington, “Teaching pronunciation: The state of the art 2021,” 2021.
- [41] K. De Bot, W. Lowie, and M. Verspoor, “A dynamic systems theory approach to second language acquisition,” *Bilingualism: Language and cognition*, vol. 10, no. 1, pp. 7–21, 2007.
- [42] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phys. Rev. A*, vol. 45, pp. 3403–3411, Mar 1992.
- [43] K. Saito and L. Plonsky, “Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis,” *Language Learning*, vol. 69, no. 3, pp. 652–708, 2019.
- [44] M. G. O’Brien, T. M. Derwing, C. Cucchiari, D. M. Hardison, H. Mixdorff, R. I. Thomson, H. Strik, J. M. Levis, M. J. Munro, J. A. Foote *et al.*, “Directions for the future of technology in pronunciation research and teaching,” *Journal of Second Language Pronunciation*, vol. 4, no. 2, pp. 182–207, 2018.